

alignment_block:

US-09-674-779-4/rev x AAY43894

Align seg 1/1 to: AAY43894 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACCATCACCACCATTAAGTCGA 11
 ||| :: ::::: ||||| ||||| ||||| ||||| |||||
 21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...ValAs 36

10 CCCTCTGCCT 1

|||||

36 pProGlyPro 39

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT: AAY43895

seq_documentation_block:

ID AAY43895 standard; Protein: 318 AA.

XX AAY43895;

XX 11-FEB-2000 (first entry)

XX Amino acid sequence of the mTNFH6NS3 clone B9 fusion protein.

KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.

XX Synthetic.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT Region 1..40

FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;

XX WPI; 2000-013283/01.

XX N-PSDB; AAZ36164.

XX New hepatitis C-virus polypeptide used for treating the infection -

XX Example 2; Fig 3-2; 66pp; English.

XX The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are also used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.

XX Sequence 318 AA;

SQ

alignment_scores:

Quality: 64.50 Length: 20

Ratio: 4.300 Gaps: 1

Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-674-779-4/rev x AAY43895

Align seg 1/1 to: AAY43895 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACCATCACCACCATTAAGTCGA 11

||| :: ::::: ||||| ||||| ||||| ||||| |||||

21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...ValAs 36

10 CCCTCTGCCT 1

|||||

36 pProGlyPro 39

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT: AAY43898

seq_documentation_block:

ID AAY43898 standard; Protein: 318 AA.

XX AAY43898;

XX 11-FEB-2000 (first entry)

XX Amino acid sequence of the mTNFH6NS3 Type 2a fusion protein.

XX HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.

XX Synthetic.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT Region 1..40

FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;

XX WPI; 2000-013283/01.

XX N-PSDB; AAZ36167.

XX New hepatitis C-virus polypeptide used for treating the infection -

XX Example 5; Fig 6-2; 66pp; English.

XX The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive

CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
 US-09-674-779-4/rev x AAY43898 ..
 Align seg 1/1 to: AAY43898 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACACATCACCATCACTTAAGTCGA 11
 ||| ::::: ::||| ||||| ||||| |||||
 21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...Valas 36

10 CCTCTGCCT 1
 |||| |||
 36 pProGlyPro 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT.AAY43899

seq_documentation_block:
 ID_ AAY43899 standard; Protein; 318 AA.
 XX AC AAY43899;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the mTNFh6NS3 Type 2b fusion protein.
 XX
 KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Region 1..40
 FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"
 XX
 PN WO9954735-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 15-APR-1999; 99WO-EP02547.
 XX
 PR 17-APR-1998; 98EP-0870087.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;
 XX
 DR WPI; 2000-013283/01.
 DR N-PSDB; AAZ36168.
 XX
 PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 6; Fig 7-2; 66pp; English.
 XX
 CC The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV

CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
 US-09-674-779-4/rev x AAY43899 ..
 Align seg 1/1 to: AAY43899 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACACATCACCATCACTTAAGTCGA 11
 ||| ::::: ::||| ||||| ||||| |||||
 21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...Valas 36

10 CCTCTGCCT 1
 |||| |||
 36 pProGlyPro 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT.AAY43900

seq_documentation_block:
 ID_ AAY43900 standard; Protein; 318 AA.
 XX AC AAY43900;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the mTNFh6NS3 Type 2c fusion protein.
 XX
 KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Region 1..40
 FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"
 XX
 PN WO9954735-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 15-APR-1999; 99WO-EP02547.
 XX
 PR 17-APR-1998; 98EP-0870087.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;
 XX
 DR WPI; 2000-013283/01.
 DR N-PSDB; AAZ36169.
 XX
 PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 7; Fig 8-2; 66pp; English.
 XX
 CC The present sequence represents a fusion protein comprising a Hepatitis C

CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-674-779-4/rev x AAY43900 ..

Align seg 1/1 to: AAY43900 from: 1 to: 318

```

60 CATGCTTACCAAAATAAGCTACACCATCACCATTAAAGTCGA 11
   |||   ::   ::   ::   ::   ::   ::   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36
   |||   |||
10 CCTCTGCCT 1
   |||   |||
36 pProGlyPro 39

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: AAY43896

seq_documentation_block:

ID AAY43896 standard; Protein; 326 AA.

XX AAY43896;

XX 11-FEB-2000 (first entry)

XX Amino acid sequence of the MTNFH6NS3 Type 3a clone 21 fusion protein.

XX HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 XX solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 XX passive vaccination.

XX Synthetic.

OS Hepatitis C virus.

XX Key Location/Qualifiers

XX Region 1..48

XX /note= "these residues represent the non-NS3
 XX sequence, which is the MTNF fusion partner,
 XX the hexahistidine tag and part of the
 XX multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;

XX WPI: 2000-013283/01.

XX N-PSDB; AA236165.

XX

PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 4; Fig 4-2; 66pp; English.
 XX

CC The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX

SQ Sequence 326 AA;

alignment_scores:

Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-674-779-4/rev x AAY43896 ..

Align seg 1/1 to: AAY43896 from: 1 to: 326

```

60 CATGCTTACCAAAATAAGCTACACCATCACCATTAAAGTCGA 11
   |||   ::   ::   ::   ::   ::   ::   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36
   |||   |||
10 CCTCTGCCT 1
   |||   |||
36 pProGlyPro 39

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: AAY43897

seq_documentation_block:

ID AAY43897 standard; Protein; 326 AA.

XX AAY43897;

XX 11-FEB-2000 (first entry)

XX Amino acid sequence of the MTNFH6NS3 Type 3a clone 32 fusion protein.

XX HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 XX solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 XX passive vaccination.

XX Synthetic.

OS Hepatitis C virus.

XX Key Location/Qualifiers

XX Region 1..48

XX /note= "these residues represent the non-NS3
 XX sequence, which is the MTNF fusion partner,
 XX the hexahistidine tag and part of the
 XX multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.

•

```

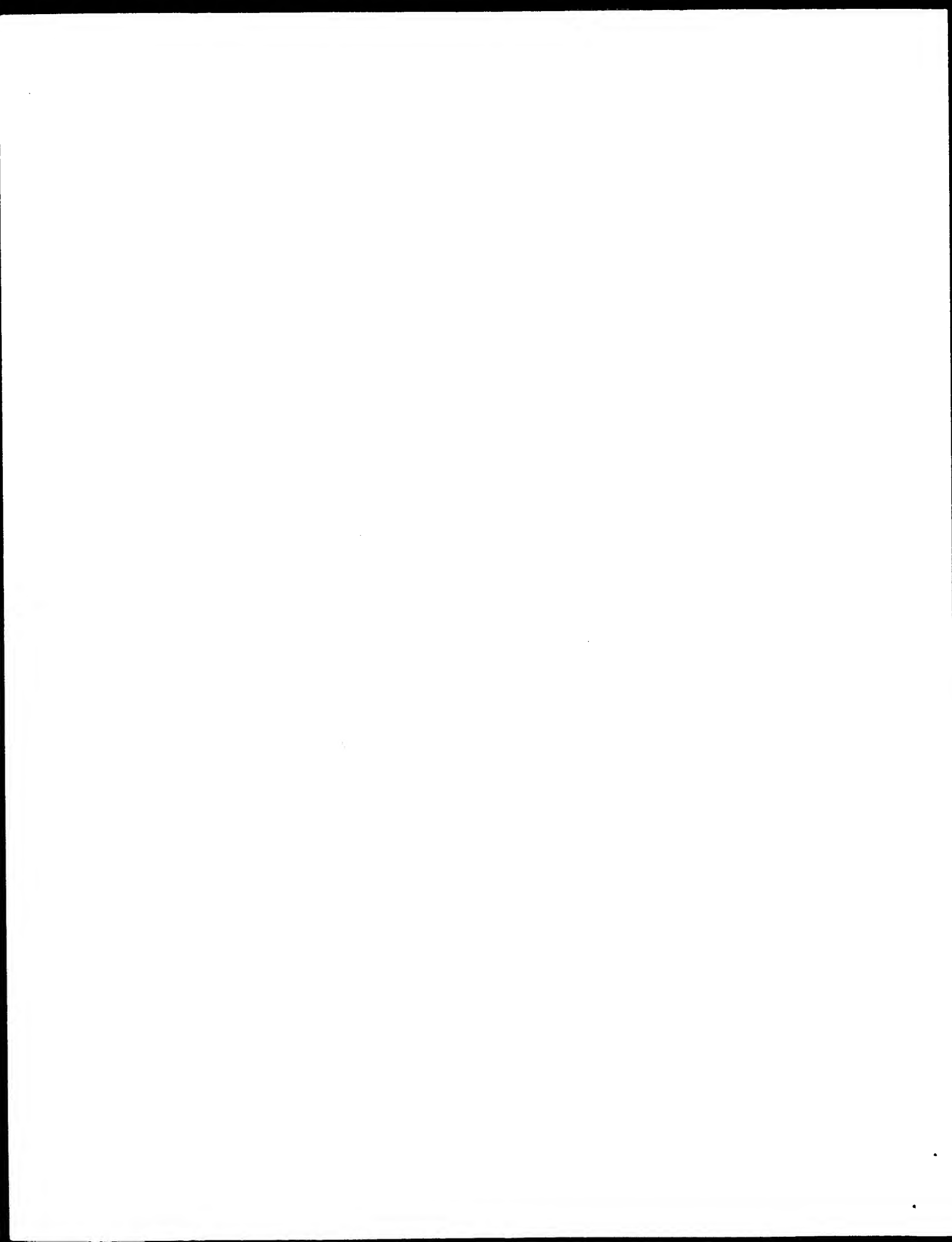
FT      /label= IL-2
FT      /note= "amino acids 1-113 of human IL-2"
FT      151
FT      /label= Spacer
FT      152..266
FT      /label= ETA
FT      /note= "amino acids 252-366 of ETA"
FT      267
FT      /label= Spacer
FT      268..413
FT      /label= GAL4
FT      /note= "amino acids 2-147 of yeast GAL4"
FT      414..421
FT      /label= Spacer
FT      /note= "endoplasmic reticulum retention signal"
XX
PN      W09613599-A1.
XX
XX      09-MAY-1996.
XX
XX      31-OCT-1995; 95WO-EP04270.
XX
XX      01-NOV-1994; 94EP-0810627.
XX
XX      (WELS/) WELS W.
XX
XX      Fominaya J, Wels W;
XX
XX      WPI; 1996-239505/24.
XX      N-PSDB; AAT29411.
XX
XX      Nucleic acid transfer system for gene therapy, e.g. against cancer
XX      - includes toxin translocation domain to target nucleic acid to
XX      specific cell
XX
XX      Claim 7; Page 67-69; 106pp; English.
XX
XX      A multidomain protein (AAR95055) has a FLAG epitope, a portion
XX      of human interleukin-2 that acts as a ligand domain, a
XX      non-cytotoxic portion of Pseudomonas aeruginosa exotoxin A acting
XX      as a translocation domain and the DNA binding domain of yeast GAL4.
XX      It is the product of a fusion gene (AAT29411) and can be expressed
XX      in E. coli (resulting in removal of an ompA signal peptide). It is
XX      used with an effector nucleic acid that comprises e.g. a gene to be
XX      delivered to a cell and a cognate structure for the GAL4 DNA binding
XX      domain. This provides a novel means of nucleic acid transfer,
XX      suitable for gene therapy.
XX
XX      Sequence 421 AA;

alignment_scores:
    Quality: 60.00      Length: 13
    Ratio: 5.455      Gaps: 0
    Percent Similarity: 84.615      Percent Identity: 69.231

alignment_block:
US-09-674-779-4/rev x AAR95055      ..
Align seg 1/1 to: AAR95055      from: 1 to: 421

45 AATAAGCTACACCATCACCATTAGTCGACCGCT 7
   :|||||
7 AspLysLeuHisHisHisHisHisHisLysLeuAlaPro 19

```




```

; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,951A
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/03499
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP/92/403403.6
; FILING DATE: 14-DEC-1992
; FILING DATE: 14-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.003A
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; FEATURE:
; NAME/KEY: mTHFMPH-taul fusion protein
; US-08-244-951A-10

alignment_scores:
  Quality: 64.50 Length: 20
  Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x US-08-244-951A-10 from: 1 to: 391
..
60 CATGCTTACCAAAATAAGCTACCATCACCATCACCATTAAAGTCGA 11
||| : : : : : : : : : : : : : : : : : : : : : : : :
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36

10 CCCTCTGCCT 1
||| |||
36 pProGlyPro 39

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-389-011-23

seq_documentation_block:
; Sequence 23, Application US/08389011
; Patent No. 5861257
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDEMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK

; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,011
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,917
; FILING DATE: 19-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,916
; FILING DATE: 19-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/244,951
; FILING DATE: 13-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/03499
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP/92/403403.6
; FILING DATE: 14-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.003-1-CON
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; FEATURE:
; NAME/KEY: mTHFMPH-taul fusion protein
; US-08-244-951A-10

alignment_scores:
  Quality: 64.50 Length: 20
  Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x US-08-389-011-23 ..
..
60 CATGCTTACCAAAATAAGCTACCATCACCATCACCATTAAAGTCGA 11
||| : : : : : : : : : : : : : : : : : : : : : : : :
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36

10 CCCTCTGCCT 1
||| |||
36 pProGlyPro 39

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.US-08-403-917A-23

seq_documentation_block:
; Sequence 23, Application US/08403917A
; Patent No. 6010913
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDEMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
```



```

/ 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840.
```

```
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB pep: US-08-849-634B-3
seq_documentation_block:
; Sequence 3, Application US/08049634B
; Patent No. 6296855
```



```

; COUNTRY: Republic of Korea
; ZIP: 706-040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,783
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 95-2751
; FILING DATE: 15-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-1000
; TELEFAX: (212) 953-7249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-600-783-13

alignment_scores:
  Quality: 58.00      Length: 11
  Ratio: 5.800       Gaps: 0
  Percent Similarity: 90.909  Percent Identity: 72.727

alignment_block:
  US-09-674-779-4/rev x US-08-600-783-13  ..

  Align seg 1/1 to: US-08-600-783-13  from: 1 to: 17

51 CCAAAATAAGCTACCATCACCACCAT 19
|||||:||||| |||||||||
1 ProSerAspLysProHisHisHisHis 11

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-677-862-2

seq_documentation_block:
; Sequence 2, Application US/08677862
; Patent No. 5874230
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,862
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J

```

```

; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-677-862-2

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708       Gaps: 1
  Percent Similarity: 52.174  Percent Identity: 43.478

alignment_block:
  US-09-674-779-4/rev x US-08-677-862-2  ..

  Align seg 1/1 to: US-08-677-862-2  from: 1 to: 763

60 CATGCTTACCAAAA.....AATAAGCT 38
||||| |||||
590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 606
||||| |||||
37 ACACCATCACCACCATCAT 19
||||| |||||
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-252-571-2

seq_documentation_block:
; Sequence 2, Application US/09252571
; Patent No. 5981250
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-674-779-4/rev x US-09-252-571-2
..
alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478
US-09-674-779-4/rev x US-09-252-571-2
..
Align seg 1/1 to: US-09-252-571-2 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-434-065-2

seq_documentation_block:
; Sequence 2, Application US/09434065
; Patent No. 6107074
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/434,065
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-434-065-2

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1

```

```

Percent Similarity: 52.174 Percent Identity: 43.478
alignment_block:
US-09-674-779-4/rev x US-09-434-065-2
..
Align seg 1/1 to: US-09-434-065-2 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-789-275-4

seq_documentation_block:
; Sequence 4, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome
; FILE REFERENCE: U 011114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 4
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;
alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478
alignment_block:
US-09-674-779-4/rev x US-08-789-275-4
..
Align seg 1/1 to: US-08-789-275-4 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-789-275-5

seq_documentation_block:
; Sequence 5, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome

```

```

; FILE REFERENCE: U 011114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 5
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Rat norvegicus
; FEATURE:
US-08-789-275-5

```

```

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
  Percent Similarity: 52.174  Percent Identity: 43.478

```

```

alignment_block:
US-09-674-779-4/rev x US-08-789-275-5 ..

```

```

Align seg 1/1 to: US-08-789-275-5 from: 1 to: 763

```

```

60 CATGCTTACCAAAA.....ATAACCT 38
||||| |||||
590 HisValAlapGInGlnAsnAlaLeuHisHisGlyAsnSerSe 606
37 ACACCATCACCACCAT 19
||||| |||||
606 rHisHisHisHisHis 612

```

```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pap:US-08-524-757-30

```

```

seq_documentation_block:
; Sequence 30, Application US/08524757
; Patent No. 5792634
; GENERAL INFORMATION:
; APPLICANT: Conaway, Ronald C.
; APPLICANT: Conaway, Joan W.
; APPLICANT: Bradsher, John N.
; TITLE OF INVENTION: RNA Polymerase Transcription Factor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13621
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160087
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Harre, John A.
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35006CIPCIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (214) 939-4500
; TELEFAX: (214) 939-4600
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-524-757-30

```

```

alignment_scores:
  Quality: 56.00      Length: 10
  Ratio: 6.222        Gaps: 0
  Percent Similarity: 90.000  Percent Identity: 80.000

```

```

alignment_block:
US-09-674-779-4/rev x US-08-524-757-30 ..

```

```

Align seg 1/1 to: US-08-524-757-30 from: 1 to: 10

```

```

39 CTACACCATCACCACCATTAAGTCGAC 10
:::||||| ||||| |||||
1 MetHisHisHisHisHisAsnValasp 10

```

```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pap:US-08-524-757-31

```

```

seq_documentation_block:
; Sequence 31, Application US/08524757
; Patent No. 5792634
; GENERAL INFORMATION:
; APPLICANT: Conaway, Ronald C.
; APPLICANT: Conaway, Joan W.
; APPLICANT: Bradsher, John N.
; TITLE OF INVENTION: RNA Polymerase Transcription Factor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13621
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160087
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Harre, John A.
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35006CIPCIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (214) 939-4500
; TELEFAX: (214) 939-4600
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-524-757-31

```

```

alignment_scores:
  Quality: 56.00      Length: 10

```

Ratio: 6.222 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x US-08-524-757-31 ..

Align seg 1/1 to: US-08-524-757-31 from: 1 to: 10

39 CTACACCATCACCATCACCATTAAAGTCGAC 10
:::|||||
1 MethHisHisHisHisHisValAsp 10

OM of: US-09-674-779-4 to: PIR_71.* out_format : pfs

Date: Jul 30, 2002 4:25 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151753_7289/app_query.fasta_1.116
-DB=PIR_71 -OPMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000 -GGAPOP=4.500
-QGAPEXT=0.050 -YGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM=ext -HEASIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09674779_@CGN1_1.113
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIME=120 -WARN_TIME=30
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: PIR_71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 28.830000
```

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
pir2:T04874	64.00	164.71	0.0416	1057	! hypothetical protein F28A21.230
pir2:S60988	62.50	167.77	0.0726	409	! hypothetical protein YOR134W -
pir1:OMRTSP	61.00	163.84	0.1276	385	! selenoprotein P precursor [Vali
pir2:S57963	58.00	156.31	0.2713	476	! methyl-CpG binding protein 2 -
pir2:A11907	59.00	156.05	0.2714	492	! methyl-CpG-binding protein 2 -
pir2:T31611	58.00	143.93	0.3989	1585	! hypothetical protein Y50F8A.g
pir2:A56235	57.00	153.79	0.5740	311	! transcription activator WafB -
pir2:T46375	57.00	151.08	0.5755	439	! hypothetical protein DKF2p434D1
pir2:T45727	56.50	154.36	0.6915	240	! hypothetical protein FIP2.220 -
pir2:JC4898	56.50	145.36	0.6975	754	! Down-syndrome-critical-region B
pir2:T32443	56.00	153.61	0.8341	219	! hypothetical protein T28B4.4 -
pir2:A38565	56.00	149.07	0.8377	390	! polycomb (PC) protein - fruit
pir2:T37451	56.00	147.63	0.8389	469	! HGF-G2 (HFG-2) protein - human
pir2:A34743	56.00	147.51	0.8390	476	! transcription factor HFK1 - hum
pir2:A43817	56.00	142.45	0.8431	906	! transforming protein (cbl) - hu
pir2:S27473	56.00	142.08	0.8434	950	! URB51 protein - smut fungus (Us
pir2:T23056	56.00	138.68	0.8462	1465	! hypothetical protein H06001.2
pir2:C86478	56.00	136.69	0.8478	1887	! protein FL504.13 [imported] -
pir2:T47447	55.50	133.58	1.02	2326	! calcium channel protein alpha -
pir2:T24446	55.00	151.08	1.21	208	! hypothetical protein T04C10.4 -
pir2:S23737	55.00	148.28	1.22	297	! prolone-rich protein precursor
pir2:T49529	55.00	147.62	1.22	323	! transcription factor-kr - mouse
pir2:S39406	55.00	146.90	1.22	354	! homeotic protein otx1 - human
pir2:S25345	55.00	146.88	1.22	355	! otx1 protein - mouse
pir2:T56547	55.00	146.88	1.22	355	! homeodomain protein otx1 - rat
pir2:A49077	55.00	146.43	1.22	378	! transcription initiation factor
pir1:A56018	55.00	145.05	1.22	448	! transcription factor Oct-6 - hu
pir1:S30205	55.00	145.03	1.22	449	! transcription factor Oct-6 - mc
pir1:A40168	55.00	145.00	1.22	451	! transcription factor Oct-6 - ra
pir2:T33997	55.00	144.66	1.22	471	! hypothetical protein W03G1.5 -
pir1:A40879	55.00	136.65	1.23	1305	! phospholipase C (EC 3.1.4.3),
pir2:T46477	55.00	136.61	1.23	1312	! phospholipase C (EC 3.1.4.3),
pir2:T47773	55.00	132.29	1.24	2273	! calcium channel BI-1 - rabbit
pir2:AA0721	54.00	141.79	1.78	467	! hypothetical protein F2A13.210
pir2:JH0672	54.00	141.67	1.78	474	! neuroblast proliferation inhibi
pir1:S31223	54.00	141.58	1.78	480	! brain factor 1 protein - rat
pir2:T17291	54.00	141.33	1.78	495	! transcription factor Brn-1 - mc
pir2:T13958	54.00	137.13	1.79	845	! hypothetical protein DKF2p434E1
pir2:T13804	54.00	134.60	1.79	1166	! synGAP-bl protein - rat
	54.00	134.30	1.79	1212	! shs protein - fruit fly (Dros

```
pir2:TL4270 - 54.00 134.06 1.79 1249 ! Ras-GTPase activating prote
pir2:TL4259 - 54.00 133.79 1.79 1293 ! ras GTPase-activating prote
pir2:T06387 - 53.50 145.20 2.14 251 ! knotted I class homeodomain
pir2:AA1098 - 53.50 128.11 2.17 2212 ! calcium channel protein alp
pir2:S37939 - 53.00 150.15 2.57 111 ! hypothetical protein YKL111c
```

seq_name: pir2:T04874

seq_documentation_block:

```
hypothetical protein F28A21.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04874
R:Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A:Accession: T04874
A:Molecule type: DNA
A:Residues: 1-1057 <BEV>
A:Cross-references: EMBL:AL035526
A:Experimental source: cultivar Columbia; BAC clone F28A21
C:Genetics:
A:Map position: 4
A:Introns: 680/3: 754/1; 881/1
A:Note: F28A21.230
```

alignment_scores:

```
Quality: 64.00 Length: 11
Ratio: 5.818 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 72.727
```

alignment_block:

```
US-09-674-779-4/rev x T04874 ..
Align seg 1/1 to: T04874 from: 1 to: 1057
51 CCAAAATAAGCTACCATCATCACCATCACCAT 19
||||:||||:||||:||||:||||:||||:||||:
928 ProARGASnVnValHisHisHisHis 938
```

seq_name: pir2:S60988

seq_documentation_block:

```
hypothetical protein YOR134W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O3320; hypothetical protein YOR33320W
C:Species: Saccharomyces cerevisiae
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 07-May-1999
C:Accession: S60988; S61690; S67019; S63865
R:Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vleck, C.; Stegemann,
submitted to the EMBL Data Library, August 1995
A:Description: Sequencing of 51 kilobases on the right arm of chromosome XV from S. c
A:Reference number: S60983
A:Accession: S60988
A:Molecule type: DNA
A:Residues: 1-409 <WTE>
A:Cross-references: EMBL:X90518; NID:g1050808; PID:g1050814
R:Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valen
submitted to the EMBL Data Library, December 1995
A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromos
A:Reference number: S61643
A:Accession: S61690
A:Molecule type: DNA
A:Residues: 1-409 <BEN>
A:Cross-references: EMBL:X94335; NID:g1262139; PID:e217752; PID:g1164978
R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansoerge, W
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66965
A:Accession: S67019
A:Molecule type: DNA
A:Residues: 1-409 <VOS>
A:Cross-references: EMBL:Z75042; NID:g1420343; PID:e252026; PID:g1420344; MIPS:YOR134
A:Experimental source: strain S288C
```

```

alignment_block:
US-09-674-779-4/rev x OMRTSP  ..

Align seg 1/1 to: OMRTSP from: 1 to: 385

54  TTACCAAAAATAAGCTACACCATCACCATCACCAT 19
|||||  :: |||||||||||||||||||
238 LeuProProSerGlyLeuHisHisHisHisHis 249

seq_name: pir2:S57963

seq_documentation_block:
methyl CpG binding protein 2 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C:Accession: S57963
R: d'Esposito, M.; Quaderi, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.
submitted to the EMBL Data Library, July 1995
A:Description: Physical mapping and expression analysis of an X-linked gene encoding
A:Reference number: S57963
A:Accession: S57963
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-476 <DES>
A:Cross-references: EMBL:X89430; NID:G899295; PIDN:CAA61599.1; PID:G899296

```

```

60 CATCTCTTACCAAAAAATAGCA...
302 HisMetLeuProArgSerGAlaLeuSerAspSerAsnAsnPheThrI 318

36 .....CACCATCACCATTAAAGTCGACCCTCGCCCT 1
318 eHisGhisHisHisHisHisAlaLeuPheProSerPro 332
```

seq_documentation_block:
 selenoprotein P precursor [validated] - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 15-Sep-2000
 C:Accession: A0380; B40380; S68322
 J:Hall, K.E.; Lloyd, R.S.; Yang, J.G.; Read, R.; Burk, R.F.
 J:Bio. Chem. 266, 10050-10053, 1991
 A:Title: The cDNA for rat selenoprotein P contains 10 TGA codons in the open
 A:Reference number: A40380; MUID:91244760
 A:Accession: A40380
 A:Molecule type: mRNA
 A:Residues: 1-385 <HIL>
 A:Cross-references: GB:M63574; NID:g206893; PIDN:AAA42129.1; PID:g206894
 A:Accession: B40380
 A:Molecule type: Protein
 A:Residues: 20-41;267-287;316-327 <HIL>
 R:Chittum, H.S.; Himeno, S.; Hill, K.E.; Burk, R.F.
 Arch. Biochem. Biophys. 325, 124-128, 1996
 A:Title: Multiple forms of selenoprotein P in rat plasma.
 A:Reference number: S68322; MUID:96140605
 A:Accession: S68322
 A:Molecule type: protein

alignment_scores:		
Quality:	61.00	Length: 12
Ratio:	6.100	Gaps: 0
Percent Similarity:	83.333	Percent Identity: 75.000

```
alignment_scores:
  Quality: 59.00      Length: 15
  Ratio: 4.917       Gaps: 0
  Percent Similarity: 80.000  Percent Identity: 60.000

alignment_block:
```


Align seg 1/1 to: T45727 from: 1 to: 240

48 AAAAATAAGTACACCATCACCATCAACCATTAAGTCGACCTCTGCCT 1
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 GluAsnGlnIleHisHisHisHisHisHisHisHisHisHisHisHis 29
.....SerProLeupro 29

seq_name: pir2:JC4898

seq_documentation_block:

Down-syndrome-critical-region protein - human
N:Alternate names: Drosophila minibrain protein homolog
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JC4898
R:Shindoh, N.; Kudoh, J.; Maeda, H.; Yamaki, A.; Minoshima, S.; Shimizu, Y.; Shimizu, N.
Biochem. Biophys. Res. Commun. 225, 92-99, 1996
A:Title: Cloning of a human homolog of the Drosophila minibrain/rat dyrk gene from "the
A:Reference number: JC4898; MUID:96332410
A:Accession: JC4898
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-754 <SHI>
A:Cross-references: DDBJ:D85759; NID:g1526445; PIDN:BAAL2866.1; PID:g1526446
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

alignment_scores:
Quality: 56.50 Length: 23
Ratio: 4.708 Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x JC4898 ..

Align seg 1/1 to: JC4898 from: 1 to: 754

60 CATGCTTACCAAAA.....AATAAGCT 38
||||| |||::: |||:::
581 HisValAlaProGlnGlnAsnAlaLeuHisHisHisHisGlyAsnSerSe 597

37 ACACCATCACCATCACCAT 19
||||| |||::: |||:::
597 rHisHisHisHisHisHisHis 603

seq_name: pir2:T32443

seq_documentation_block:

hypothetical protein T28B4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T32443
R:Wilson, R.; Greco, T.; Sansone, J.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T28B4.
A:Reference number: Z21168
A:Accession: T32443
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-219 <WIL>
A:Cross-references: EMBL:AF026206; PIDN:AAB71260.1; GSPDB:GN00028; CESP:T28B4.4
A:Experimental source: strain Bristol N2; clone T28B4
C:Genetics:
A:Gene: CESP:T28B4.4
A:Map position: X
A:Introns: 34/2; 138/2; 184/3

alignment_scores:
Quality: 56.00 Length: 9
Ratio: 6.222 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:

US-09-674-779-4/rev x T32443 ..

Align seg 1/1 to: T32443 from: 1 to: 219

45 AATAAGCTACACCATCACCATCACCAT 19
:::|||||:|||||:|||||:|||||:|||||
35 HisLysMetHisHisHisHisHisHis 43

seq_name: pir1:A38565

seq_documentation_block:

polycomb (Pc) protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38565
R:Paro, R.; Hogness, D.S.
Proc. Natl. Acad. Sci. U.S.A. 88, 263-267, 1991
A:Title: The polycomb protein shares a homologous domain with a heterochromatin-associ-

A:Reference number: A38565; MUID:91095442

A:Accession: A38565

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-390 <PAR>

A:Cross-references: GB:X55702; NID:g8321; PIDN:CAA392229.1; PID:g603986

C:Genetics:

A:Gene: FlyBase:Pc

A:Cross-references: FlyBase:FBgn0003042

C:Superfamily: polycomb protein; chromobox homolog

F:26-63/Domain: chromobox homolog <CBH>

alignment_scores:

Quality: 56.00 Length: 10
Ratio: 6.222 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x A38565 ..

Align seg 1/1 to: A38565 from: 1 to: 390

48 AAAATAAGCTACACCATCACCATCACCAT 19
|||::: ||| |||::: |||::: |||:::
130 LysGluLysLysHisHisHisHisHisHis 139

seq_name: pir2:I37451

seq_documentation_block:

HBF-G2 (HFK-2) protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 24-Sep-1999
C:Accession: I37451
R:Wiese, S.; Murphy, D.B.; Schlung, A.; Burfeind, P.; Schmundt, D.; Schnulle, V.; Mat
Biochim. Biophys. Acta 1262, 105-112, 1995
A:Title: The genes for human brain factor 1 and 2, members of the fork head gene fami
A:Reference number: I37451; MUID:95322450
A:Accession: I37451
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-469 <RES>
A:Cross-references: EMBL:X78202; NID:g967047; PIDN:CAA55038.1; PID:g967048
C:Genetics:
A:Gene: HBF-G2; HFK-2
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homolog
F:162-253/Domain: fork head DNA-binding domain homolog <FHD>

alignment_scores:
Quality: 56.00 Length: 17
Ratio: 5.091 Gaps: 0
Percent Similarity: 64.706 Percent Identity: 52.941

OM of: US-09-674-779-4 to: SwissProt_40:* out_format : pfs

Date: Jul 30, 2002 4:31 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (C) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cqn2.1/USPTO.spool/US09674779/runat_30072002_151754_7376/app_query.fasta_1.116
-DB=SwissProt_40 -OFAST=fastan -SURFIX-rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGNL_1_45 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-4

Query length: 60

Database: SwissProt_40:*

Database sequences: 105224

Database length: 38719550

Search time (sec): 16.440000

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
SwissProt_40:SELP_RAT	61.00	159.40	0.0909	385	1	P25236 rattus norvegicus (rat)
SwissProt_40:MBC2_HUMAN	59.00	151.91	0.1882	486	1	P51608 homo sapiens (human)
SwissProt_40:MBC2_RAT	59.00	151.81	0.1882	492	1	Q00568 rattus norvegicus (rat)
SwissProt_40:MBC2_MOUSE	58.00	149.12	0.2704	484	1	Q922d6 mus musculus (mouse)
SwissProt_40:SNF1_CANGA	57.00	144.45	0.3996	611	1	Q00372 candida glabrata (yea)
SwissProt_40:DYRA_HUMAN	56.50	141.29	0.4682	763	1	Q13627 homo sapiens (human)
SwissProt_40:DYRA_MOUSE	56.50	141.29	0.4682	763	1	Q61214 mus musculus (mouse)
SwissProt_40:DYRA_RAT	56.50	141.29	0.4682	763	1	Q63470 rattus norvegicus (rat)
SwissProt_40:PC_DROME	56.00	145.17	0.5567	390	1	P26017 drosophila melanogaster
SwissProt_40:FXGA_HUMAN	56.00	143.72	0.5580	469	1	P55316 homo sapiens (human)
SwissProt_40:FXGB_HUMAN	56.00	143.58	0.5581	477	1	P55315 homo sapiens (human)
SwissProt_40:GEL_HUMAN	56.00	138.52	0.5624	906	1	P22681 homo sapiens (human)
SwissProt_40:URB1_USTMA	56.00	138.15	0.5627	950	1	P40349 ustilago maydis (smut)
SwissProt_40:CCAB_DIOSM	55.50	129.67	0.6818	2326	1	P56698 discopyge ommata (el)
SwissProt_40:MAF1_MOUSE	55.00	143.83	0.7981	323	1	P54841 mus musculus (mouse)
SwissProt_40:MAF1_RAT	55.00	143.83	0.7981	323	1	P54842 rattus norvegicus (rat)
SwissProt_40:OTX1_HUMAN	55.00	143.11	0.7989	354	1	P32242 homo sapiens (human)
SwissProt_40:OTX1_MOUSE	55.00	143.09	0.7990	355	1	P80205 mus musculus (mouse)
SwissProt_40:OTX1_RAT	55.00	143.09	0.7990	355	1	P80205 mus musculus (mouse)
SwissProt_40:72AA_HUMAN	55.00	142.63	0.7995	355	1	Q63410 rattus norvegicus (rat)
SwissProt_40:OCT6_HUMAN	55.00	141.25	0.8012	448	1	P52655 homo sapiens (human)
SwissProt_40:OCT6_MOUSE	55.00	141.23	0.8012	448	1	Q03052 homo sapiens (human)
SwissProt_40:OCT6_RAT	55.00	141.20	0.8013	451	1	P21952 mus musculus (mouse)
SwissProt_40:PIPI_DROME	55.00	132.77	0.8116	1312	1	P20267 rattus norvegicus (rat)
SwissProt_40:C258_HUMAN	54.00	143.38	1.114	239	1	P54555 drosophila melanogaster
SwissProt_40:ANA_DROME	54.00	137.98	1.15	474	1	P58505 homo sapiens (human)
SwissProt_40:FXGB_RAT	54.00	137.88	1.15	480	1	Q26307 drosophila melanogaster
SwissProt_40:FXGB_MOUSE	54.00	137.87	1.15	481	1	Q00939 rattus norvegicus (rat)
SwissProt_40:ONC2_HUMAN	54.00	137.80	1.15	481	1	Q06987 mus musculus (mouse)
SwissProt_40:BRN1_MOUSE	54.00	137.64	1.15	485	1	Q09548 homo sapiens (human)
SwissProt_40:BRN1_RAT	54.00	137.61	1.15	495	1	P31361 mus musculus (mouse)
SwissProt_40:BRN1_HUMAN	54.00	137.61	1.15	497	1	Q63262 rattus norvegicus (rat)
SwissProt_40:CBXA_HUMAN	54.00	137.56	1.15	500	1	P20264 homo sapiens (human)
SwissProt_40:ECR_LUCCU	54.00	136.69	1.15	558	1	P20264 homo sapiens (human)
SwissProt_40:PHYB_SORBI	54.00	134.29	1.16	757	1	Q00257 mus musculus (human)
SwissProt_40:BUN2_DROME	54.00	130.80	1.17	1178	1	Q18531 lucilia cuprina (green)
SwissProt_40:CCAA_HUMAN	54.00	130.58	1.17	1211	1	P91527 sorghum bicolor (sorgh)
SwissProt_40:CCAA_MOUSE	54.00	124.84	1.18	2505	1	Q24523 drosophila melanogaster
SwissProt_40:CCAA_RAT	53.50	124.59	1.41	2164	1	Q00555 homo sapiens (human)
SwissProt_40:YKLI_YEAST	53.50	124.41	1.41	2212	1	P97445 mus musculus (mouse)
SwissProt_40:YKLI_MOUSE	53.00	146.61	1.63	111	1	P54282 rattus norvegicus (rat)
SwissProt_40:YKLI_RAT	53.00	146.61	1.63	111	1	P36074 saccharomyces cerevis

SwissProt_40:CAR2_DICDI - 53.00 137.00 1.65 375 ! P34907 dictyostelium disc
SwissProt_40:SELP_MOUSE - 53.00 136.90 1.65 380 ! P70274 mus musculus (mous
SwissProt_40:FXB2_MOUSE - 53.00 135.96 1.65 428 ! Q64733 mus musculus (mous
SwissProt_40:PDML_DROME - 53.00 133.28 1.66 601 ! P31368 drosophila melanog
SwissProt_40:YG51_YEAST - 53.00 131.17 1.67 785 ! P50089 saccharomyces cere

seq_name: SwissProt_40:SELP_RAT

seq_documentation_block:

ID SELP_RAT STANDARD; PRT; 385 AA.
AC P25236;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Selenoprotein P precursor (Sep).
GN SEPL1 OR SELP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC TISSUE=Liver;
RX MEDLINE=91244760; PubMed=2037562;
RA Hill K.E., Lloyd R.S., Yang J.-G., Read R., Burk R.F.;
RT "The cDNA for rat selenoprotein P contains 10 TGA codons in the open
reading frame.";
RL J. Biol. Chem. 266:10050-10053(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Kidney;
RX MEDLINE=95364621; PubMed=7637580;
RA Sailoh K., Saito N., Lee M.J., Fujii M., Kobayashi T., Sumino K.;
RT "Molecular cloning of cDNA encoding a bovine selenoprotein P-like
protein containing 12 selenocysteines and a (His-Pro) rich domain
insertion, and its regional expression.";
RL Brain Res. Mol. Brain Res. 30:301-311(1995).
RN [3]
RP REVIEW.
RX MEDLINE=95017128; PubMed=7931697;
RA Burk R.F., Hill K.E.;
RT "Selenoprotein P, A selenium-rich extracellular glycoprotein.";
RC J. Nutr. 124:1891-1897(1994).
CC -! FUNCTION: MIGHT BE RESPONSIBLE FOR SOME OF THE EXTRACELLULAR
CC THE ANTIOXIDANT DEFENSE PROPERTIES OF SELENIUM OR MIGHT BE INVOLVED IN
CC THE TRANSPORT OF SELENIUM.
CC -! SURCELLULAR LOCATION: Extracellular.
CC -! TISSUE SPECIFICITY: MADE IN THE LIVER AND SECRETED INTO THE
CC PLASMA.
CC -! MISCELLANEOUS: THE SELENOCYSTEINES ARE ALL ENCODED BY THE OPAL
CC CODON, UGA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M63574; AAA42129.1; -
EMBL; D25221; BAA04950.2; -
PIR; A40380; OMRTSP.
KW Glycoprotein; Signal; Selenium; Selenocysteine; Plasma.
FT SIGNAL 1 19
FT CHAIN 20 385 SELENOPROTEIN P.
FT DOMAIN 244 252 POLY-HIS.
FT SE_CYS 59 59
FT SE_CYS 264 264
FT SE_CYS 282 282
FT SE_CYS 323 323
FT SE_CYS 335 335
FT SE_CYS 357 357

```

FT SE_CYS 371 371
FT SE_CYS 373 373
FT SE_CYS 380 380
FT SE_CYS 382 382
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 385 AA: 42614 MW; 013HDF4FB741E3E8 CRC64;

alignment_scores:
  Quality: 61.00 Length: 12
  Ratio: 6.100 Gaps: 0
  Percent Similarity: 83.333 Percent Identity: 75.000

alignment_block:
US-09-674-779-4/rev x SEMP_RAT ..
Align seg 1/1 to: SEMP_RAT from: 1 to: 385
54 TTACCAAAAATAAGCTACACATCACATCATCAT 19
||||| :||| ||||| ||||| |||||
238 LeuProSerGlyLeuHisHisHisHisHis 249

seq_name: SwissProt_40:MECP2_HUMAN

seq_documentation_block:
ID MECP2_HUMAN STANDARD; PRT; 486 AA.
AC P51608; O15233;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN MECP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBTAXID:9606;
[1]
RP SEQUENCE FROM N.A.
RA Kudo S., Fukuda M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RA Thiesen J., Straetling W.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE-97130625; PubMed-8976388;
RA Vilain A., Apollon F., Vogt N., Dutrillaux B., Malfroy B.;
RT "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization."
RL Cytogenet. Cell Genet. 74:293-294 (1996).
[4]
RP SEQUENCE FROM N.A.
RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MeCP2 locus."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RX MEDLINE-95299240; PubMed-10369871;
RA Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression."
RL Hum. Mol. Genet. 8:1253-1262 (1999).
[6]
RP SEQUENCE OF 10-486 FROM N.A.
RT TISSUE-Skeletal muscle;

```

```

RX MEDLINE-96327611; PubMed-8672133;
RA D'Esposito M., Quaderi N.A., Ciccodicola A., Bruni P., Esposito T.,
RA D'Urso M., Brown S.D.M.;
RT "Isolation, physical mapping, and Northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2."
RL Mamm. Genome 7:533-535 (1996).
[7]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA Sandoval N., Rosenthal A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[8]
RP VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX MEDLINE-99438392; PubMed-10508514;
RA Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RA Zoghbi H.Y.;
RT "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2."
RL Nat. Genet. 23:185-188 (1999).
[9]
RP VARIANT RTT VAL-140.
RX MEDLINE-20465115; PubMed-11007980;
RA Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
RA Poon P.M., Zappella M., Federico A., Sorrentino V.;
RT "MECP2 mutation in male patients with non-specific X-linked mental retardation."
RL FEBS Lett. 481:285-288 (2000).
[10]
RP VARIANTS RTT W-106; F-124; C-13; C-134; R-152; M-158 AND C-306.
RX MEDLINE-20439334; PubMed-10991688;
RA Obata K., Matsuishi T., Yamashita Y., Fukuda T., Kuwajima K.,
RA Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
RA Mori K., Kondo I.;
RT "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in patients with Rett syndrome."
RL J. Med. Genet. 37:608-610 (2000).
[11]
RP VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
RX MEDLINE-20439335; PubMed-10991689;
RA Hampson K., Woods C.G., Latif F., Webb T.;
RT "Mutations in the MECP2 gene in a cohort of girls with Rett syndrome."
RL J. Med. Genet. 37:610-612 (2000).
[12]
RP VARIANT RTT VAL-140.
RX PubMed-11805248;
RA Dotti M.T., Orrico A., De Stefano N., Battisti C., Sicurelli F.,
RA Severi S., Lam C.W., Galli L., Sorrentino V., Federico A.;
RT "A Rett syndrome MECP2 mutation that causes mental retardation in men."
RL Neurology 58:226-230 (2002).
-!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN BIND SPECIFICALLY TO A SINGLE METHYL-CpG PAIR. IT IS NOT INFLUENCED BY SEQUENCES FLANKING THE METHYL-CpGS. MEDIATES TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE DEACETYLASE AND THE COREPRESSOR SIN3A.
-!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CpG IN THE GENOME.
-!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
-!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL 6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA, INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS. AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS USUALLY SURVIVE INTO ADULTHOOD.
-!- SIMILARITY: CONTAINS (R)GRE(K) MOTIFS AND SPKK MOTIFS THAT HAVE BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
-!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).

```


AC Q92206:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Methyl-Cpg-binding protein 2 (Mecp2-2 protein) (Mecp2).
GN MECP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT Cpg binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99299240; PubMed=10369871;
RA Coy J.F., Sedlacek Z., Baechner D., Delli H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative
RT polyadenylation within the long 3'-untranslated region of the methyl-
RT Cpg-binding protein 2 gene (Mecp2) suggests a regulatory role in gene
RT expression.";
RL Hum. Mol. Genet. 8:1253-1262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Reichwald K., Thiessen J., Wiehe T., Kioschis P., Straetling W.H.,
RX Rosenthal A., Platzer M.;
RT "Comparative analysis of the methyl Cpg binding protein 2 locus in man
RT and mouse reveals new untranslated sequences.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
CC BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
CC INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGs. MEDIATES
CC TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
CC DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
CC GENOME.
CC -1- SIMILARITY: CONTAINS (R)G(RP)(K) MOTIFS AND SPKK MOTIFS THAT HAVE
CC BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
CC -1- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF072251; AAC68880.1; -
DR EMBL; AJ132522; CAB46495.1; -
DR EMBL; AF121351; AAF22116.1; -
DR EMBL; AF158181; AAF33024.1; -
DR MGD; MGI:99918; Mecp2.
DR InterPro: IPR000637; At_hook.
DR InterPro: IPR001739; MBD.
DR Pfam; PF01429; MBD; 1.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00391; MBD; 1.
DR Transcription regulation; Repressor; DNA-binding; Nuclear protein.
KW DOMAIN 96 149 MBD.
FT DOMAIN 277 283 POLY-ALA.
FT DOMAIN 366 372 POLY-HIS.
FT DOMAIN 384 391 POLY-PRO.
FT DOMAIN 440 443 POLY-THR.
SQ SEQUENCE 484 AA; 52307 MW; 62FD228F0118A49F CRC64;

alignment_scores:
Quality: 57.00 Length: 10

seq_name: SwissProt_40:SNF1_CANGA
seq_documentation_block:
ID SNF1_CANGA STANDARD; PRT; 611 AA.
AC Q00372;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Carbon catabolite derepressing protein kinase (EC 2.7.1.1).
GN SNF1.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCLS84;
RX MEDLINE=97101049; PubMed=8945576;
RA Petter R., Kwon-Chung K.J.;
RT "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT pathogenic yeast Candida glabrata.";
RL Infect. Immun. 64:5269-5273(1996).
CC -1- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT
CC INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY
CC PROTEIN SNF4. COULD PHOSPHORYLATES CAT8 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC SNF1 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L78130; AAB48642.1; -
DR HSSP; P00518; 2PHK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT DOMAIN 6 17 POLY-HIS.
FT DOMAIN 39 290 PROTEIN KINASE.
FT NP_BIND 45 53 ATP (BY SIMILARITY).
FT BINDING 68 68 ATP (BY SIMILARITY).
FT ACT_SITE 161 161 BY SIMILARITY.
FT MOD_RES 194 194 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 611 AA; 70049 MW; 89E17812A4900CD0 CRC64;

alignment_scores:
Quality: 57.00 Length: 10

Ratio: 6.333 Caps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x SNF1_CANGA ..

Align seg 1/1 to: SNF1_CANGA from: 1 to: 611

48 AAAAATAAGTACACCATCACCACCATCACCACCAT 19

||||| ||||||| ||||||| ||||||| |||||||

2 GluAsnLysGluHisHisHisHisHisHisHis 11

seq_name: SwissProt_40: DYRA_HUMAN

seq_documentation_block:

ID DYRA_HUMAN STANDARD; PRT: 763 RA.

AC Q13627; Q92810; Q92582; Q9UNW5;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A

(EC 2.7.1.-) (Protein kinase minibrain homolog) (MNBH) (HP86) (Dual

specificity YAK1-related kinase).

GN DYRK1A OR DYRK OR MNBH OR MNB.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (LONG ISOFORM).

RX MEDLINE=97131512; PubMed=8975710;

RA Song W.J., Sternberg L.R., Kasten-Sportes C., van Keuren M.L.,

RA Chung S.H., Slack A.C., Miller D.E., Glover T.W., Chiang P.W., Lou L.,

RA Kurnit D.W.;

RT "Isolation of human and murine homologues of the Drosophila minibrain

RT gene: human homologue maps to 21q22.2 in the Down syndrome 'critical

RT region'.";

RL Genomics 38:331-339(1996).

RN [2]

RP SEQUENCE FROM N.A. (LONG ISOFORM).

RX MEDLINE=97026291; PubMed=8872470;

RA Guimera A.M., Casas C., Pucharcos C., Solans A., Domenech A.,

RA Planas A.M., Ashley J., Lovett M., Estivill X., Pritchard M.A.;

RT "A human homologue of Drosophila minibrain (MNB) is expressed in the

RT neuronal regions affected in Down syndrome and maps to the critical

RT region.";

RL Hum. Mol. Genet. 5:1305-1310(1996).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RX TISSUE=Fetal brain;

RA MEDLINE=96332410; PubMed=8769099;

RA Shindoh N., Kudoh J., Maeda H., Yamaki A., Minoshima S., Shimizu Y.,

RA Shimizu N.;

RT "Cloning of a human homolog of the Drosophila minibrain/rat Dyck gene

RT from 'the Down syndrome critical region' of chromosome 21.";

RL Biochem. Biophys. Res. Commun. 225:92-99(1996).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Miki O., Masahira H., Yasufumi M., Naohiko S., Takahiro N.,

RA Hitoshi I., Nobuo N., Toshihiko E., Yoshiyuki S., Misao O.,

RA Ohira M., Seki N., Nagase T., Suzuki E., Nomura N., Ohara O.,

RA Hattori M., Sakaki Y., Eki T., Murakami Y., Saito T., Ichikawa H.,

RA Ohki M.;

RT Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A., VARIANTS F-415; H-681, AND ALTERNATIVE SPLICING.

RX MEDLINE=99263503; PubMed=10329007;

RA Guimera J., Casas C., Estivill X., Pritchard M.;

RT "Human minibrain homologue (MNBH/DYRK1): characterization, alternative

RT splicing, differential tissue expression, and overexpression in Down

RT syndrome.";

RL Genomics 57:407-418(1999).

RN CC

-!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING

NUCLEAR FUNCTIONS OF CELL PROLIFERATION. PHOSPHORYLATES SERINES,
THREONINES AND TYROSINES RESIDUES IN ITS SEQUENCE AND IN EXOGENOUS
SUBSTRATES.

-!- SUBCELLULAR LOCATION: Nuclear.

-!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; A LONG FORM (SHOWN
HERE), 1, 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.

-!- TISSUE SPECIFICITY: UBIQUITOUS. HIGHEST LEVELS IN SKELETAL MUSCLE,

TESTIS, FETAL LUNG AND FETAL KIDNEY.

-!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING CENTRAL NERVOUS
SYSTEM.

-!- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.

-!- DISEASE: OVEREXPRESSED 1.5-FOLD IN FETAL DOWN SYNDROME BRAIN.

-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

MNB/DYRK SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U58496; AAC50939.1; -

EMBL; U52373; AAB18639.1; -

EMBL; D85759; BAAL2866.1; -

EMBL; D86550; BAAL3110.1; -

EMBL; AF108830; AAD31169.1; -

HSSP; P27703; IERK.

MIM; 600855; -

InterPro: IPR000719; Euk_pkinase.

InterPro: IPR002290; Ser_thr_pkinase.

Pfam: PF00069; pkinase; 2.

SMART; SM00220; S_TKc; 1.

PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;

ATP-binding; Nuclear protein; Phosphorylation; Alternative splicing;

Polymorphism.

DOMAIN 117 134

DOMAIN 159 173

NP_BIND 165 479

PROTEIN KINASE

ATP (BY SIMILARITY).

BINDING 188 188

ATP (BY SIMILARITY).

ACT_SITE 287 287

BY SIMILARITY.

DOMAIN 509 515

POLY-SER.

DOMAIN 599 602

POLY-HIS.

DOMAIN 607 619

SER/THR-RICH.

DOMAIN 656 672

POLY-SER.

MOD_RES 219 219

PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

MOD_RES 319 319

PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

MOD_RES 321 321

PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

VARSPPLIC 70 78

MISSING (IN ISOFORM 1).

VARSPPLIC 516 529

GGSGTSGNGRARS -> GASAISSSWLYRH (IN

ISOFORM 2).

VARSPPLIC 530 763

MISSING (IN ISOFORM 2).

VARSPPLIC 516 540

GGSGTSGNGRARDPPTHQHRHSGG -> VEQHWMPGAFRM

TVSFTEVHDVVPV (IN ISOFORM 3).

MISSING (IN ISOFORM 3).

VARSPPLIC 541 763

ROQFPAPLWGSQTAPQTVVETHPV -> SSHVVHLLVSP

AILEWSSTGCQVPLE (IN ISOFORM 4).

VARSPPLIC 559 584

MISSING (IN ISOFORM 4).

VARSPPLIC 585 763

Y -> F.

VARSPPLIC 415 415

/FTID=VAR_009395.

Q -> H.

VARSPPLIC 681 681

/FTID=VAR_009396.

G -> A (IN REF. 1).

CONFLICT 32 32

N -> S (IN REF. 1).

CONFLICT 47 47

S -> P (IN REF. 1).

CONFLICT 57 57

S -> R (IN REF. 1).

CONFLICT 123 123

K -> N (IN REF. 1).

CONFLICT 397 397

FT CONFLICT 592 592 A -> G (IN REF. 1).
 FT CONFLICT 679 679 A -> P (IN REF. 1).
 SQ SEQUENCE 763 AA; 85584 MW; 7C3A52A3CB04FB5 CRC64;

alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x DYRA_HUMAN ..

Align seg 1/1 to: DYRA_HUMAN from: 1 to: 763

60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||||
 590 HisValalaProGlnGlnAsnAlaLeuHisHisHisHisGlyAsnSerSe 606

37 ACACCATCACCACCAT 19
 ||||| ||||| ||||| |||||

606 rHisHisHisHisHisHis 612

seq_name: SwissProt_40:DYRA_MOUSE

seq_documentation_block:

ID DYRA_MOUSE STANDARD; PRT; 763 AA.
 AC Q61214;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A
 DE (EC 2.7.1.1) (Protein kinase minibrain homolog) (MNBH) (MP86) (Dual
 DE specificity YAK1-related kinase).
 GN DYRK1A OR DYRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER / NIH;
 RX MEDLINE=97131512; PubMed=8975710;
 RA Song W.J., Sternberg L.R., Kasten-Sportes C., van Keuren M.L.,
 RA Chung S.H., Slack A.C., Miller D.E., Glover T.W., Chiang P.W., Lou L.,
 RA Kurnit D.W.;
 RA "Isolation of human and murine homologues of the Drosophila minibrain
 RT gene: human homologue maps to 21q22.2 in the Down syndrome 'critical
 RT region'";
 RL Genomics 38:331-339(1996).
 RN [2]
 RP SUBCELLULAR LOCATION, AND SUBUNIT.
 RX MEDLINE=97224401; PubMed=9070862;
 RA Song W.J., Chung S.H., Kurnit D.M.;
 RT "The murine Dyk protein maps to chromosome 16, localizes to the
 RT nucleus, and can form multimers";
 RL Biochem. Biophys. Res. Commun. 231:640-644(1997).
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING
 CC NUCLEAR FUNCTIONS OF CELL PROLIFERATION, PHOSPHORYLATES SERINES,
 CC THREONINES AND TYROSINES RESIDUES IN ITS SEQUENCE AND IN EXOGENOUS
 CC SUBSTRATES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MNB/DYRK SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U58497; AAC52994.1; ..
 DR HSPP; P27703; IERK
 DR MGD; MGI-1330299; Dyrla.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 2.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_SF; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Nuclear protein; Phosphorylation.
 FT BIPARTITE NUCLEAR LOCALIZATION SIGNAL
 FT DOMAIN 117 134
 FT PROTEIN KINASE.
 FT NP_BIND 159 479
 FT NP_BIND 165 173
 FT BINDING 188 188
 FT ACT_SITE 287 287
 FT DOMAIN 509 515
 FT DOMAIN 599 602
 FT DOMAIN 607 619
 FT DOMAIN 656 672
 FT DOMAIN 664 671
 FT MOD_RES 219 219
 FT MOD_RES 319 319
 FT MOD_RES 321 321
 SQ SEQUENCE 763 AA; 85494 MW; E117DDD6C5E8C74F CRC64;

alignment_scores:

Quality: 56.50 Length: 23
 Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x DYRA_MOUSE ..

Align seg 1/1 to: DYRA_MOUSE from: 1 to: 763

60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||||
 590 HisValalaProGlnGlnAsnAlaLeuHisHisHisHisGlyAsnSerSe 606

37 ACACCATCACCACCAT 19
 ||||| ||||| ||||| |||||

606 rHisHisHisHisHisHis 612

seq_name: SwissProt_40:DYRA_MOUSE

seq_documentation_block:

ID DYRA_MOUSE STANDARD; PRT; 763 AA.
 AC Q63470;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A
 DE (EC 2.7.1.1) (Protein kinase minibrain homolog) (MNBH) (RP86) (Dual
 DE specificity YAK1-related kinase).
 GN DYRK1A OR DYRK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A., AND MUTAGENESIS OF TYR-219.
 RP STRAIN=SPRAGUE-DAWLEY; Tissue=Brain;
 RX MEDLINE=96216443; PubMed=8631952;
 RA Kenrup H., Becker W., Heukelbach J., Wilmes A., Schuermann A.,
 RA Hupertz C., Raulainen H., Joost H.-G.;
 RA "Dyrk, a dual specificity protein kinase with unique structural
 RT features whose activity is dependent on tyrosine residues between
 RT subdomains VII and VIII.";

RL J. Biol. Chem. 271:3488-3495(1996).
 RN [2]
 RP REVISIONS.
 RA Kentrup H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE-98421512; PubMed-9748265;
 RA Becker W., Weber Y., Wetzel K., Eilmbter K., Tejedor F.J.,
 RA Joost H.-G.;
 RT "Sequence characteristics, subcellular localization, and substrate
 RT specificity of DYRK-related kinases, a novel family of dual
 RT specificity protein kinases.";
 RL J. Biol. Chem. 273:25893-25902(1998).
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING
 CC NUCLEAR FUNCTIONS OF CELL PROLIFERATION. PHOSPHORYLATES SERINES,
 CC THREONINES AND TYROSINES RESIDUES IN ITS SEQUENCE AND IN EXOGENOUS
 CC SUBSTRATES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -!- PFM: AUTOPHOSPHORYLATED ON TYR RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MNE/DYRK SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; X79769; CAA56164.1; --
 DR HSPSP; P27703; IERK.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 2.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_SP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Nuclear protein; BIPARTITE NUCLEAR LOCALIZATION SIGNAL
 FT DOMAIN 117 134
 FT 159 479
 FT NP_BIND 165 173
 FT BINDING 188 188
 FT ACT_SITE 287 287
 FT DOMAIN 509 515
 FT 599 602
 FT DOMAIN 607 619
 FT 656 672
 FT DOMAIN 664 671
 FT 219 219
 FT MOD_RES 319 319
 FT MOD_RES 321 321
 FT MOD_RES 70 78
 FT VARSPLIC 70 78
 FT MUTAGEN 219 219
 FT Y->F: REDUCED AUTOPHOSPHORYLATION ON
 FT TYROSINE, BUT NO LOSS OF HISTONE
 FT PHOSPHORYLATION.
 FT SEQUENCE 763 AA; 85541 MW; CB5EC7EC4C1F9A47 CRC64;

 alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

 alignment_block:
 US-09-674-779-4/rev x DYRA_RAT ..

Align seg 1/1 to: DYRA_RAT from: 1 to: 763
 60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||||
 590 HsValAlaProGlnGlnAsnAlaLeuHsHisHsHisGlyAsnSer 606
 ||||| |||||
 37 ACACATCACCATCACCAT 19
 ||||| |||||
 606 rHisHisHisHisHisHis 612
 seq_name: SwissProt_40:PC_DROME
 seq_documentation_block:
 ID PC_DROME STANDARD; PRT; 390 AA.
 AC P26017; Q9VP49;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Polycarb protein.
 GN PC OR C67618.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE-91095442; PubMed-1898775;
 RA Paro R., Hogness D.S.;
 RT "The Polycarb protein shares a homologous domain with a
 RT heterochromatin-associated protein of Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:263-267(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch L., Beasley E.M.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolshakov S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
 RA Borkova D., Botchan M.R., Bouck J., Brinkley P., Cantrell D., Chandra I.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

FT DOMAIN 124 127 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN_FING 381 420 RING-TYPE.
FT DOMAIN 357 476 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 477 688 PRO-RICH.
FT DOMAIN 689 834 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 856 895 UBA.
FT MOD_RES 700 700 PHOSPHORYLATION.
FT MOD_RES 774 774 PHOSPHORYLATION.
SQ SEQUENCE 906 AA; 99646 MW; 7D686B050204AD8F CRC64;

alignment_scores:
Quality: 56.00 Length: 12
Ratio: 5.600 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:

US-09-674-779-4/rev x CBL_HUMAN ..

Align seg 1/1 to: CBL_HUMAN from: 1 to: 906

36 CACCATCACCATCACCATTAAAGTCGACCCCTCGCT 1

||||| : : : : : |||

36 HishishishishishishisLeuSerProHisPro 47

seq_name: SwissProt_40:URBL_USTMA

seq_documentation_block:

ID URBL_USTMA STANDARD; PRT; 950 AA.

AC P40349;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Siderophore biosynthesis regulatory protein URBS1.

GN URBS1.

OS Ustilago maydis (Smut fungus).

OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;

OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.

OX NCBI_TaxID=5270;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=518;

RX MEDLINE=94019380; PubMed=8413298;

RA Volzard C.P., Wang J., McEvoy J.L., Xu P., Leong S.A.;

RT "urbs1, a gene regulating siderophore biosynthesis in Ustilago

RT maydis, encodes a protein similar to the erythroid transcription

RT factor GATA-1.";

RL Mol. Cell. Biol. 13:7091-7100(1993).

CC -!- FUNCTION: INVOLVED IN THE REGULATION OF SECRETED PERRICHROME-TYPE

CC SIDEROPHORES. ACTS DIRECTLY OR INDIRECTLY TO REPRESS THE

CC BIOSYNTHESIS OF SIDEROPHORES.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M80547; AAB05617.1; -.

DR PIR; S27473; S27473.

DR HSSP; P17679; 1GNF.

DR TRANSFAC; T02406; -.

DR InterPro; IPR000679; Znf_GATA.

DR InterPro; IPR001164; Znf_GCS.

DR Pfam; PF00320; GATA; 2.

DR PRINTS; PR00619; GATAZNFINGER.

DR SMART; SM00401; ZNF_GATA; 2.

DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.

DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.

KW DNA-binding; Zinc-finger; Transcription regulation; Repressor;
KW Nuclear protein.
FT ZN_FING 338 362 GATA-TYPE 1.
FT ZN_FING 482 506 GATA-TYPE 2.
FT DOMAIN 24 27 POLY-ALA.
FT DOMAIN 28 36 POLY-SER.
FT DOMAIN 487 490 POLY-THR.
FT DOMAIN 753 780 HIS-RICH.
FT DOMAIN 902 906 POLY-SER.
SQ SEQUENCE 950 AA; 101427 MW; F969C6DA09A78C12 CRC64;

alignment_scores:
Quality: 56.00 Length: 13
Ratio: 5.600 Gaps: 0
Percent Similarity: 76.923 Percent Identity: 61.538

alignment_block:

US-09-674-779-4/rev x URBL_USTMA ..

Align seg 1/1 to: URBL_USTMA from: 1 to: 950

57 GCTTACCAAAATAAGCTACCATCACCATCACCAT 19

||||| : : : : : |||

750 ValValProHishishishishishishishisHis 762

seq_name: SwissProt_40:CCAB_DISOM

seq_documentation_block:

ID CCAB_DISOM STANDARD; PRT; 2326 AA.

AC P56698;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Probable voltage-dependent N-type calcium channel alpha-1B subunit

DE (DOE-4).

OS Discopyle ommata (Electric ray).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Squala; Hypnosqualea; Pristiogalea; Batoidae;

OC Torpediniformes; Narcinoidei; Narcinidae; Discopyge.

OX NCBI_TaxID=7785;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Electric lobe;

RX MEDLINE=93248175; PubMed=7683405;

RA Horne W.A., Ellinor P.T., Inman I., Zhou M., Tsien R.W., Schwarz T.L.;

RT "Molecular diversity of Ca²⁺ channel alpha 1 subunits from the marine

RT ray Discopyle ommata.";

RL Proc. Natl. Acad. Sci. U.S.A. 90:3787-3791(1993).

CC -!- FUNCTION: THE ISOFORM ALPHA-1B GIVES RISE TO N-TYPE CALCIUM

CC CURRENTS. N-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE

CC ACTIVATED" (HVA) GROUP (BY SIMILARITY).

CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT

CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS

CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-

CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS

CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM

CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA

CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY

CC SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).

CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY

CC ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN THE ELECTRIC LOBE THAN

CC IN THE FOREBRAIN.

CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE

CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE

CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS

CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A

CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.

CC -!- PTM: PHOSPHORYLATED IN VITRO BY CAM-KINASE II, CAPK, PKC AND CGPK

CC (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS

FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L12532; NOT_ANNOTATED_CDS.

InterPro: IPR002077; Ca_channel.

InterPro: IPR002111; Cat_channel_trpl.

InterPro: IPR000636; Cation_chan_non_lig.

InterPro: IPR001682; Channel_pore_Ca_Na.

Pfam: PF00520; Ion_trans; 4.

PRINTS: PR00167; CACHANNEL.

KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;

KW Calcium channel; Glycoprotein; Repeat; Multigene family;

KW Calcium-binding; Phosphorylation; Alternative splicing.

FT REPEAT 75 351 I.

FT REPEAT 458 702 II.

FT REPEAT 1134 1416 III.

FT REPEAT 1453 1708 IV.

FT DOMAIN 1 88 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 89 107 S1 OF REPEAT I (POTENTIAL).

FT DOMAIN 108 125 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 126 145 S2 OF REPEAT I (POTENTIAL).

FT DOMAIN 146 156 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 157 176 S3 OF REPEAT I (POTENTIAL).

FT DOMAIN 177 180 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 181 199 S4 OF REPEAT I (POTENTIAL).

FT DOMAIN 200 218 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 219 238 S5 OF REPEAT I (POTENTIAL).

FT DOMAIN 239 323 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 324 348 S6 OF REPEAT I (POTENTIAL).

FT DOMAIN 349 472 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 473 491 S1 OF REPEAT II (POTENTIAL).

FT DOMAIN 492 506 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 507 526 S2 OF REPEAT II (POTENTIAL).

FT DOMAIN 527 534 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 535 552 S3 OF REPEAT II (POTENTIAL).

FT DOMAIN 553 563 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 564 582 S4 OF REPEAT II (POTENTIAL).

FT DOMAIN 583 601 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 602 621 S5 OF REPEAT II (POTENTIAL).

FT DOMAIN 622 674 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 675 699 S6 OF REPEAT II (POTENTIAL).

FT DOMAIN 700 1148 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1149 1166 S1 OF REPEAT III (POTENTIAL).

FT DOMAIN 1167 1182 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1183 1202 S2 OF REPEAT III (POTENTIAL).

FT DOMAIN 1203 1214 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1215 1233 S3 OF REPEAT III (POTENTIAL).

FT DOMAIN 1234 1243 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1244 1262 S4 OF REPEAT III (POTENTIAL).

FT DOMAIN 1263 1281 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1282 1301 S5 OF REPEAT III (POTENTIAL).

FT DOMAIN 1302 1388 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1389 1413 S6 OF REPEAT III (POTENTIAL).

FT DOMAIN 1414 1468 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1469 1487 S1 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1488 1502 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1503 1522 S2 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1523 1530 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1531 1549 S3 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1550 1558 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1559 1577 S4 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1578 1596 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1597 1616 S5 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1617 1680 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1681 1705 S6 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1706 2326 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 1869 1873 POLY-GLN.

FT DOMAIN 2040 2046 POLY-HIS.

FT DOMAIN 371 398 BINDING TO THE BETA SUBUNIT (BY SIMILARITY).

FT SITE 306 306 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 653 653 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 1362 1362 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 1650 1650 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT MOD_RES 1716 1716 PHOSPHORYLATION (BY CAPK) (POTENTIAL).

FT CA_BIND 1734 1745 BY SIMILARITY.

FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 406 406 D -> DDGLGIIIEPEQKPEDIQSVY (IN ISOPFORM 2).

SEQ SEQUENCE 2326 AA; 264515 MW; D5DEAA09E819B6B CRC64;

alignment_scores:

Quality: 55.50 Length: 23

Ratio: 4.625 Gaps: 1

Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x CCAB_DISOM ..

Align seg 1/1 to: CCAB_DISOM from: 1 to: 2326

60 CATGCTCTTACCAAAAATAGCTA..... 37

2023 HisValMetProAspTyrSerLeuGluArgValleProValGlnMetPr 2039

36 .CACCATCACCATCACCAT 19

2039 OHHisHisHisHisHis 2045

seq_name: SwissProt_40:MAFL_MOUSE

seq_documentation_block:

ID MAFL_MOUSE STANDARD; PRT; 323 AA.

AC P54841.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription factor MAFL (Segmentation protein KR) (Kreisler).

GN MAFB OR MAFL OR KRML.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95094266; PubMed=8001130;

RA Cordes S.P., Barsh G.S.;

RT "The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.";

RL Cell 79:1025-1034(1994).

CC -!- FUNCTION: MAY PLAY AN EARLY ROLE IN AXIAL PATTERNING (HINDBRAIN SEGMENTATION).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.

CC -!- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND

CC IN THE CAUDAL HINDRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),

CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT

CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY

CC THROUGH R6.

CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L36435; AAA65689.1; -.
DR HSSP; P05412; 1JUN.
DR TRANSFAC; T01439; -.
DR MGD; MGI:104555; MafB.
DR InterPro; IPR001871; bZIP.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 131 143 POLY-HIS.
FT DOMAIN 158 167 POLY-HIS.
FT DNA_BIND 238 264 BASIC MOTIF.
FT DOMAIN 266 287 LEUCINE-ZIPPER.
FT MUTAGEN 248 248 N->S: LOSS OF TRANSCRIPTIONAL ACTIVITY.
SQ SEQUENCE 323 AA; 35809 MW; D77AE07ABD9C2AD2 CRC64;

alignment_scores:
Quality: 55.00 Length: 12
Ratio: 5.500 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:

US-09-674-779-4/rev x MAF1_MOUSE ..

Align seg 1/1 to: MAF1_MOUSE from: 1 to: 323

36 CACCATCACCATCACCATTAGTCGACCCCTCTGCT 1
|||||
162 HISHISHISHISHISGlnAlaSerProProPro 173

OM of: US-09-674-779-4 to: SPREMBL_19:* out_format : pfs

Date: Jul 30, 2002 4:31 PM

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```

NAME=frame+inp_model
-Q=/cgn2.1/USPNO_spool/US09674779/runat_30072002_151754_7349/app_query.fasta_1.116
-DB=CGN2MEMBL19 -OFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEAT=4.000 -MINMATCH=0.100 -LOOPCT=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEAT=0.050 -XGAPOP=10.000 -XGAPEAT=0.500
-FGAPOP=6.000 -FGAPEAT=7.000 -XGAPOP=10.000 -YGAPEAT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS=human40.cgi -LIST=45 -DALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674779.@CN1_1_205 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XIPXY -WAIT -THREADS=1

```

Search information block:

Query: US-09-674-779-4

Query length: 60

Database: SPTREMBL 19:**

Database sequences: 562222

Database length: 172994929

Search time (sec): 47.140000

score_list:	Strd	Orig	zscore	EScore	Len	Document
ssp_plant:Q9SN33	-	64.00	168.26	0.0475	1057	Q9SN33 arabidopsis thaliana (md)
ssp_fungi:Q12128	-	62.50	171.53	0.0807	409	Q12128 saccharomyces cerevisiae (yeast)
ssp_invertebrate:Q24746	-	62.00	165.08	0.1010	747	Q24746 drosophila virilis (fruit fly)
ssp_plant:Q9J5L8	-	59.00	170.90	0.2934	122	Q9J5L8 macaca fascicularis (macaque)
ssp_mammal:Q95LC8	-	59.00	159.56	0.3151	486	Q95LC8 macaca fascicularis (macaque)
ssp_invertebrate:Q9VZ60	-	59.00	132.56	0.3291	1128	Q9VZ60 drosophila melanogaster (fruit fly)
ssp_invertebrate:Q9VZ58	-	59.00	151.71	0.3310	1266	Q9VZ58 plasmodium falciparum (malaria parasite)
ssp_human:Q9GF26	-	58.00	157.11	0.4621	454	Q9GF26 homo sapiens (human)
ssp_invertebrate:Q9YGC6	-	58.00	156.88	0.4628	467	Q9YGC6 xenopus laevis (african clawed toad)
ssp_human:Q9HCK8	-	58.00	147.78	0.4901	1417	Q9HCK8 homo sapiens (human)
ssp_invertebrate:Q9YCM3	-	58.00	144.37	0.5007	2146	Q9YCM3 gallus gallus (chicken)
ssp_invertebrate:Q9PW44	-	58.00	144.33	0.5008	2157	Q9PW44 gallus gallus (chicken)
ssp_invertebrate:Q9PW46	-	58.00	144.28	0.5010	2171	Q9PW46 gallus gallus (chicken)
ssp_invertebrate:Q9PMU5	-	58.00	144.24	0.5011	2182	Q9PMU5 gallus gallus (chicken)
ssp_invertebrate:Q9PMU4	-	58.00	143.73	0.5027	2321	Q9PMU4 gallus gallus (chicken)
ssp_invertebrate:Q9PW45	-	58.00	143.69	0.5028	2332	Q9PW45 gallus gallus (chicken)
ssp_invertebrate:Q9PW47	-	58.00	143.64	0.5030	2346	Q9PW47 gallus gallus (chicken)
ssp_invertebrate:Q9PMU6	-	58.00	143.61	0.5031	2357	Q9PMU6 gallus gallus (chicken)
ssp_plant:Q94F76	-	57.00	160.94	0.6515	1957	Q94F76 zea mays (maize)
ssp_invertebrate:Q9SN320	-	57.00	158.79	0.6603	256	Q9SN320 caenorhabditis elegans (nematode)
ssp_invertebrate:Q9N888	-	57.00	157.20	0.6670	311	Q9N888 gallus gallus (chicken)
ssp_invertebrate:Q90370	-	57.00	157.20	0.6670	311	Q90370 coturnix coturnix japonica (quail)
ssp_human:Q9NSV0	-	57.00	154.37	0.6790	439	Q9NSV0 homo sapiens (human)
ssp_invertebrate:Q9I863	-	57.00	154.30	0.6793	443	Q9I863 fuqu rubripes (japanese quail)
ssp_invertebrate:Q9SG87	-	57.00	150.74	0.6947	684	Q9SG87 arabidopsis thaliana (md)
ssp_plant:Q9SGP1	-	57.00	149.62	0.6996	784	Q9SGP1 calliphora vicina (blow fly)
ssp_invertebrate:Q9SN71	-	56.50	157.82	0.7995	240	Q9SN71 arabidopsis thaliana (md)
ssp_invertebrate:Q9VUE1	-	56.00	163.32	0.9269	102	Q9VUE1 drosophila melanogaster (fruit fly)
ssp_invertebrate:Q17346	-	56.00	157.06	0.9641	219	Q17346 caenorhabditis elegans (nematode)
ssp_invertebrate:Q9N890	-	56.00	154.00	0.9829	318	Q9N890 trypanosoma brucei (african trypanosome)
ssp_invertebrate:Q9SVY5	-	56.00	149.71	1.01	537	Q9SVY5 anopheles gambiae (african mosquito)
ssp_invertebrate:Q24742	-	56.00	148.74	1.02	604	Q24742 xenopus laevis (african clawed toad)
ssp_invertebrate:Q9VXX9	-	56.00	148.29	1.02	638	Q9VXX9 drosophila melanogaster (fruit fly)
ssp_invertebrate:Q9VED3	-	56.00	143.21	1.05	1186	Q9VED3 drosophila melanogaster (fruit fly)
ssp_invertebrate:Q17909	-	56.00	141.48	1.06	1465	Q17909 caenorhabditis elegans (nematode)
ssp_plant:Q9LQH2	-	56.00	139.40	1.08	1887	Q9LQH2 arabidopsis thaliana (md)
ssp_invertebrate:Q9V745	-	55.50	143.00	1.27	1012	Q9V745 drosophila melanogaster (fruit fly)
ssp_invertebrate:Q22156	-	55.00	154.47	1.42	208	Q22156 caenorhabditis elegans (nematode)
ssp_human:Q9P282	-	55.00	153.19	1.43	243	Q9P282 homo sapiens (human)
ssp_mammal:Q9N2G3	-	55.00	153.19	1.43	243	Q9N2G3 pan troglodytes (chimpanzee)


```

ID Q95LG8 PRELIMINARY; PRT; 486 AA.
AC Q95LG8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE METHYL-CFG-BINDING PROTEIN 2.
GN MECP2.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Muramatsu S.;
RT "Excessive hand-wringing in a mmp2-treated monkey.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF295597; AAK97131.1; -.
SQ SEQUENCE 486 AA; 52426 MW; 3471B61D90D92A7D CRC64;

alignment_scores:
  Quality: 59.00 Length: 15
  Ratio: 4.917 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 60.000

alignment_block:
US-09-674-779-4/rev x Q95LG8 ..
Align seg 1/1 to: Q95LG8 from: 1 to: 486

51 CCAAAATAAGTACACCATCACCACCATCAGTTCGACCCCT 7
|||||:|||||:|||||:|||||:|||||:|||||:
362 ProLysLysGluHisHisHisSerGluSerPro 376

seq_name: sp_invertebrate:Q9VZ60

seq_documentation_block:
ID Q9VZ60 PRELIMINARY; PRT; 1128 AA.
AC Q9VZ60;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG2186 PROTEIN.
GN CG2186.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Arl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47967.1; -.
DR FlyBase; FBgn0030243; CG2186.
SQ SEQUENCE 1128 AA; 122705 MW; EA6FA5F44B8D645C CRC64;

alignment_scores:
  Quality: 59.00 Length: 13
  Ratio: 5.364 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 61.538

alignment_block:
US-09-674-779-4/rev x Q9VZ60 ..
Align seg 1/1 to: Q9VZ60 from: 1 to: 1128

60 CATGCTTTACCAAAATAAGTACACCATCACCACCATCAC 22
|||||:|||||:|||||:|||||:|||||:|||||:
910 HisLeuAlaProSerSerLysGlnHisHisHisHis 922

seq_name: sp_invertebrate:Q95Z58

seq_documentation_block:
ID Q95Z58 PRELIMINARY; PRT; 1266 AA.
AC Q95Z58;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KRUEPPEL-LIKE PROTEIN.
GN KRUEPPEL.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=3D7;
RX Boschet C.J., Vaquero C.M.;
RT "Plasmodium falciparum krueppel-like gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ318337; CAC42230.1; -.
SQ SEQUENCE 1266 AA; 151646 MW; BBD631DE99DEFC8 CRC64;

alignment_scores:
  Quality: 59.00 Length: 16
  Ratio: 4.917 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
US-09-674-779-4/rev x Q95Z58 ..
Align seg 1/1 to: Q95Z58 from: 1 to: 1266

48 AAAATAAGTACACCATCACCACCATCAGTTCGACCCCT 1
|||||:|||||:|||||:|||||:|||||:|||||:

```

506 LysAsnAsnLysHisHisHisHisHisHisLeuLysTyrIlePro 521

```
seq_name: sp_human:Q96F26
seq_documentation_block:
ID Q96F26 PRELIMINARY; PRT; 454 AA.
AC Q96F26;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3534143) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
FL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011695; AAL11695.1; -.
FT NON_TER
SQ SEQUENCE 454 AA; 50269 MW; 11FC856EF3EEF24E CRC64;

alignment_scores:
Quality: 58.00 Length: 14
Ratio: 5.273 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:
US-09-674-779-4/rev x Q96F26 ..
Align seg 1/1 to: Q96F26 from: 1 to: 454

60 CATGCTTACCAAAATAAGCTACACATCACCATCACCAT 19
||||| :|:|:| :|:|:|:|:|:|:|:|:|:|:|:|
358 HisValAspSerThrMetLeuHisHisHisHisHis 371

seq_name: sp_vertebrate:Q9YGC6
seq_documentation_block:
ID Q9YGC6 PRELIMINARY; PRT; 467 AA.
AC Q9YGC6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE METHYL-CPG-BINDING PROTEIN MECP2.
GN MECP2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98282101; PubMed=9620779;
RA Jones P.L., Veenstra G.J.C., Wade P.A., Vermaak D., Kass S.U.,
RA Landsberger N., Strouboulis J., Wolffe A.P.;
RT "Methylated DNA and MeCP2 recruit histone deacetylase to repress
transcription.";
RL Nat. Genet. 19:187-191(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kass S.U., Strouboulis J., Wolffe A.P.;
RT "Xenopus MeCP2.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF106951; AAD03736.1; -.
DR EMBL; AF051768; AAD02651.1; -.
DR InterPro; IPR001739; MBD.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 467 AA; 51757 MW; 5D3A719A59E560BC CRC64;
```

```
alignment_scores:
Quality: 58.00 Length: 14
Ratio: 4.833 Gaps: 1
Percent Similarity: 85.714 Percent Identity: 71.429

alignment_block:
US-09-674-779-4/rev x Q9YGC6 ..
Align seg 1/1 to: Q9YGC6 from: 1 to: 467

54 TTACCAAAATAAGCTA.....CACATCACCATCACCAT 19
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
356 LeuProLysLysGluLeuGlnGlnHisHisHisHisHis 369

seq_name: sp_human:Q9HCK8
seq_documentation_block:
ID Q9HCK8 PRELIMINARY; PRT; 1417 AA.
AC Q9HCK8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1564 PROTEIN (FRAGMENT).
GN KIAA1564.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046784; BAB13390.1; -.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
KW ATP-binding; Helicase.
FT NON_TER
SQ SEQUENCE 1417 AA; 161268 MW; 199854F63D8BC6AE CRC64;

alignment_scores:
Quality: 58.00 Length: 14
Ratio: 5.273 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:
US-09-674-779-4/rev x Q9HCK8 ..
Align seg 1/1 to: Q9HCK8 from: 1 to: 1417

60 CATGCTTACCAAAATAAGCTACACATCACCATCACCAT 19
||||| :|:|:| :|:|:|:|:|:|:|:|:|:|:|:|
1321 HisValAspSerThrMetLeuHisHisHisHisHis 1334

seq_name: sp_vertebrate:Q9PUM3
seq_documentation_block:
ID Q9PUM3 PRELIMINARY; PRT; 2146 AA.
AC Q9PUM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB4 VARIANT.
GN CAC41B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=DORSAL ROOT GANGLION NEURONS;
 RX MEDLINE=20044702; PubMed=10574919;
 RA Lu Q., Dunlap K.;
 RT "Cloning and functional expression of novel N-type Ca(2+) channel
 variants.";
 RL J. Biol. Chem. 274:34566-34575(1999).
 DR EMBL; AF173015; AAD51818.1; -
 DR InterPro; IPR000636; Cation_chan_non_lig.
 DR InterPro; IPR002111; Cat_channel_TripL.
 DR InterPro; IPR002077; Ca_channel.
 DR InterPro; IPR001682; Channel_pore_Ca_Na.
 DR Pfam; PF00520; ion_trans; 4.
 DR PRINTS; PR00167; CACHANNEL.
 DR PRINTS; PR00170; NACHANNEL.
 SQ SEQUENCE 2146 AA; 244650 MW; 281BECAC49AE49 CRC64;

alignment_scores:
 Quality: 58.00 Length: 12
 Ratio: 5.273 Gaps: 0
 Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
 US-09-674-779-4/rev x Q9PUM3 ..
 Align seg 1/1 to: Q9PUM3 from: 1 to: 2146
 54 TTACCAAAATAAGCTACACCATCATCATCAT 19
 2036 ValProThrAspGlnValHisHisHisHisHis 2047

seq_name: sp Vertebrate:Q9PW44

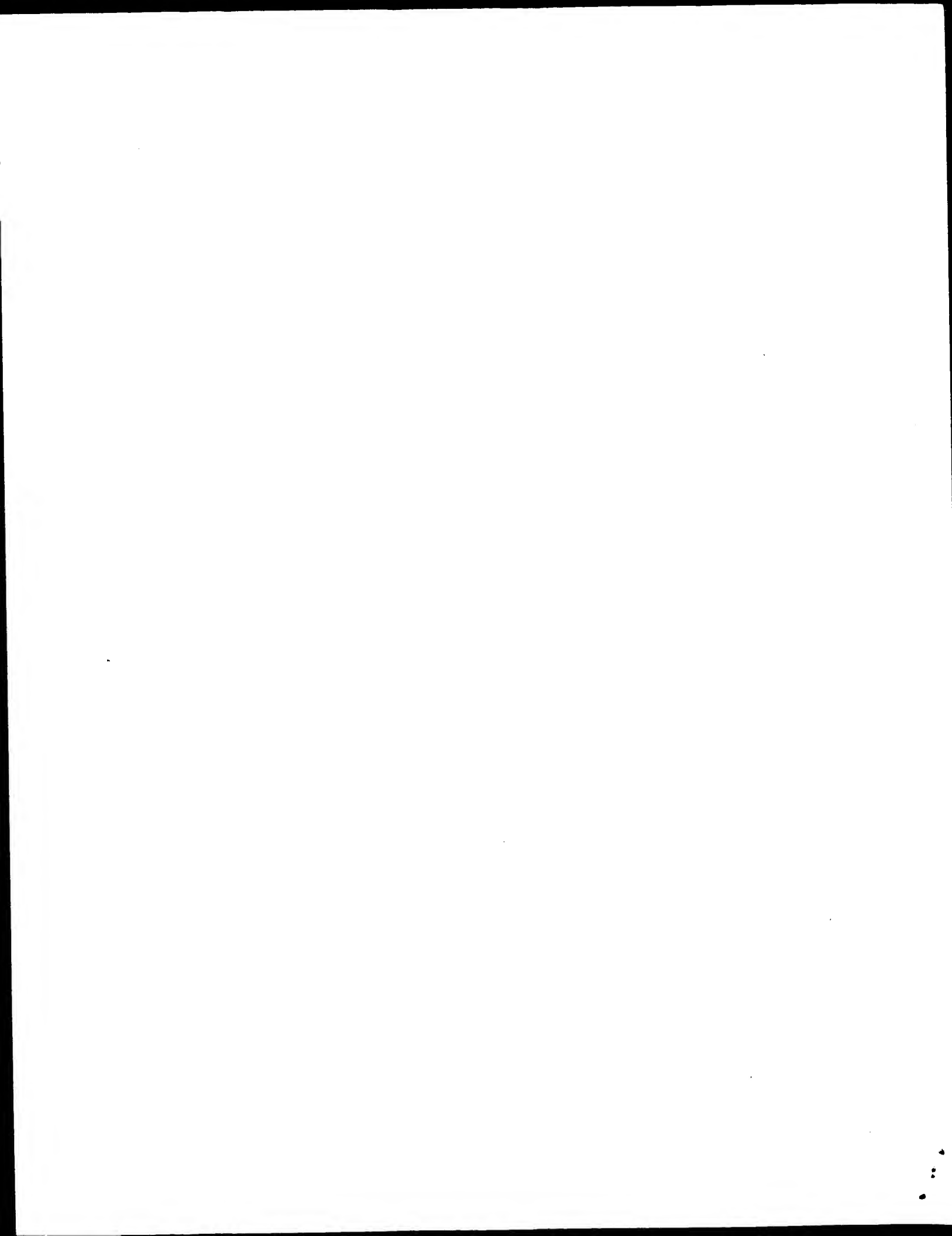
seq_documentation_block:
 ID Q9PW44 PRELIMINARY; PRT; 2157 AA.
 AC Q9PW44;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE N-YPE CALCIUM CHANNEL ALPHA-1B CDB8 VARIANT.
 GN CAC4A1B.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=DORSAL ROOT GANGLION NEURONS;
 RA Lu Q., Dunlap K.;
 RT "Cloning and Functional Expression of Novel N-type Ca2+ Channel
 Variants.";
 RL J. Biol. Chem. 0:0-0(1999).
 DR EMBL; AF173019; AAD51822.1; -
 DR InterPro; IPR000636; Cation_chan_non_lig.
 DR InterPro; IPR002111; Cat_channel_TripL.
 DR InterPro; IPR002077; Ca_channel.
 DR InterPro; IPR001682; Channel_pore_Ca_Na.
 DR Pfam; PF00520; ion_trans; 4.
 DR PRINTS; PR00167; CACHANNEL.
 DR PRINTS; PR00170; NACHANNEL.
 SQ SEQUENCE 2157 AA; 245854 MW; 5E85C3AAF9493589 CRC64;

alignment_scores:
 Quality: 58.00 Length: 12

Wed Jul 31 08:20:12 2002

us-09-674-779-4.rspt

Page 7



OM of: US-09-674-779-4 to: A_Geneseq_032802:* out_format : pfs

Date: Jul 30, 2002 4:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151835_8501/app_query.fasta_1.116
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=oligo.rag
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200
-THR_SCORE=quality -THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779@cgnl_1_88 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-4
Query length: 60
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 53.240000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig zScore EScore Len ! Documentation ..



OM of: US-09-674-779-4 to: Issued_Patents_AA:* out_format : pfs

Date: Jul 30, 2002 4:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet.n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151836_8512/app_query.fasta_1.116
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=oligo.ra1
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200
-THR_SCORE=quality -THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_29 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-674-779-4

Query length: 60

Database: Issued_Patents_AA:*

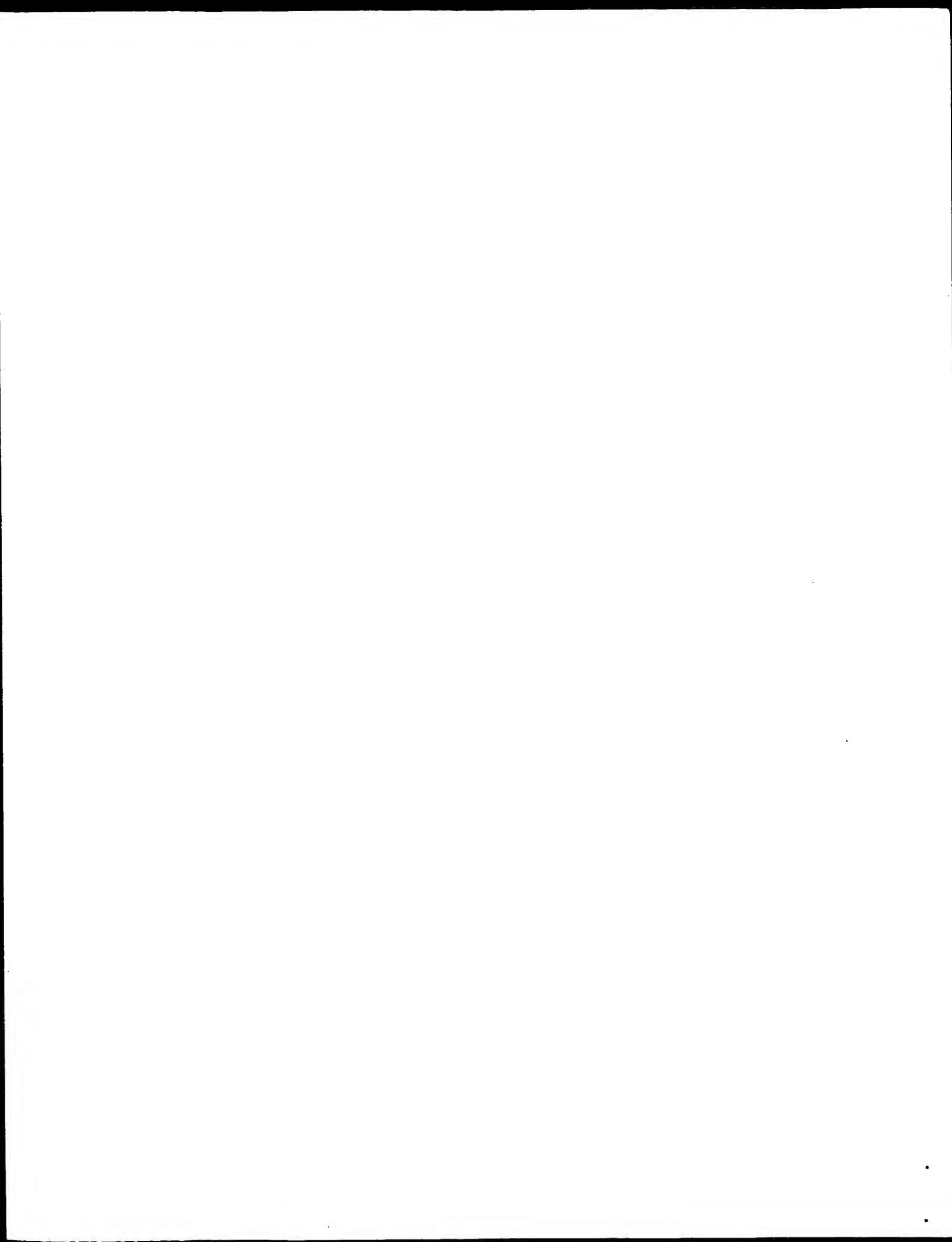
Database sequences: 231628

Database length: 24425594

Search time (sec): 22.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:	Strd Orig	zScore	EScore	Len	! Documentation	..
Sequence						



OM of: US-09-674-779-4 to: PIR_71:* out_format : pfs

Date: Jul 30, 2002 4:33 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09674779/runat_30072002_151836_8525/app_query.fasta_1.116
-DB=PIR_71 -QFMT=fastan -SUFFIX=oligo.rpr -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779.ecgn1_1.113 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

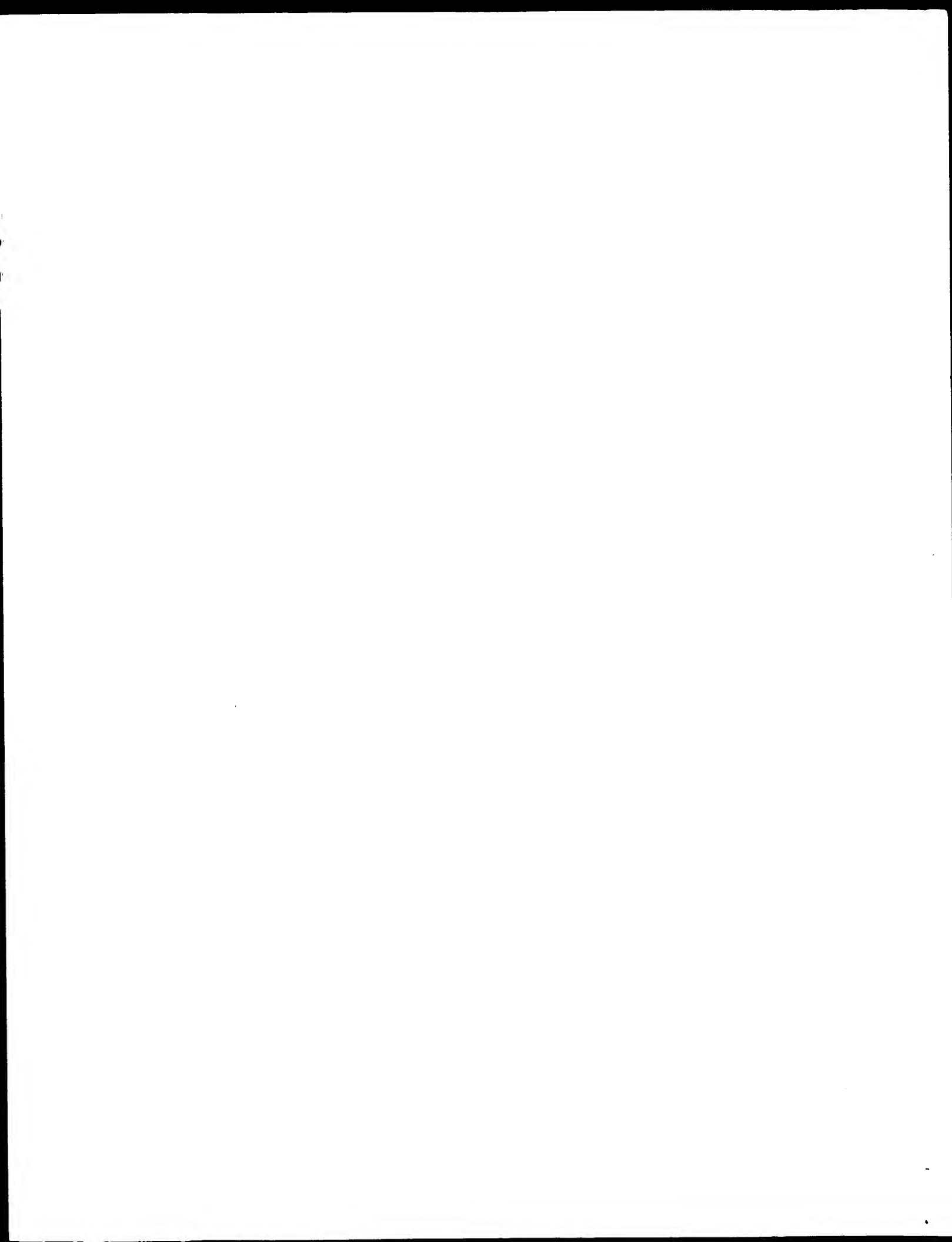
Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 28.890000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig zScore EScore Len ! Documentation ..



OM of: US-09-674-779-4 to: SwissProt_40:* out_format : pfs

Date: Jul 30, 2002 4:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151837_8635/app_query.fasta_1.116
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=oligo.rsp -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_45 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

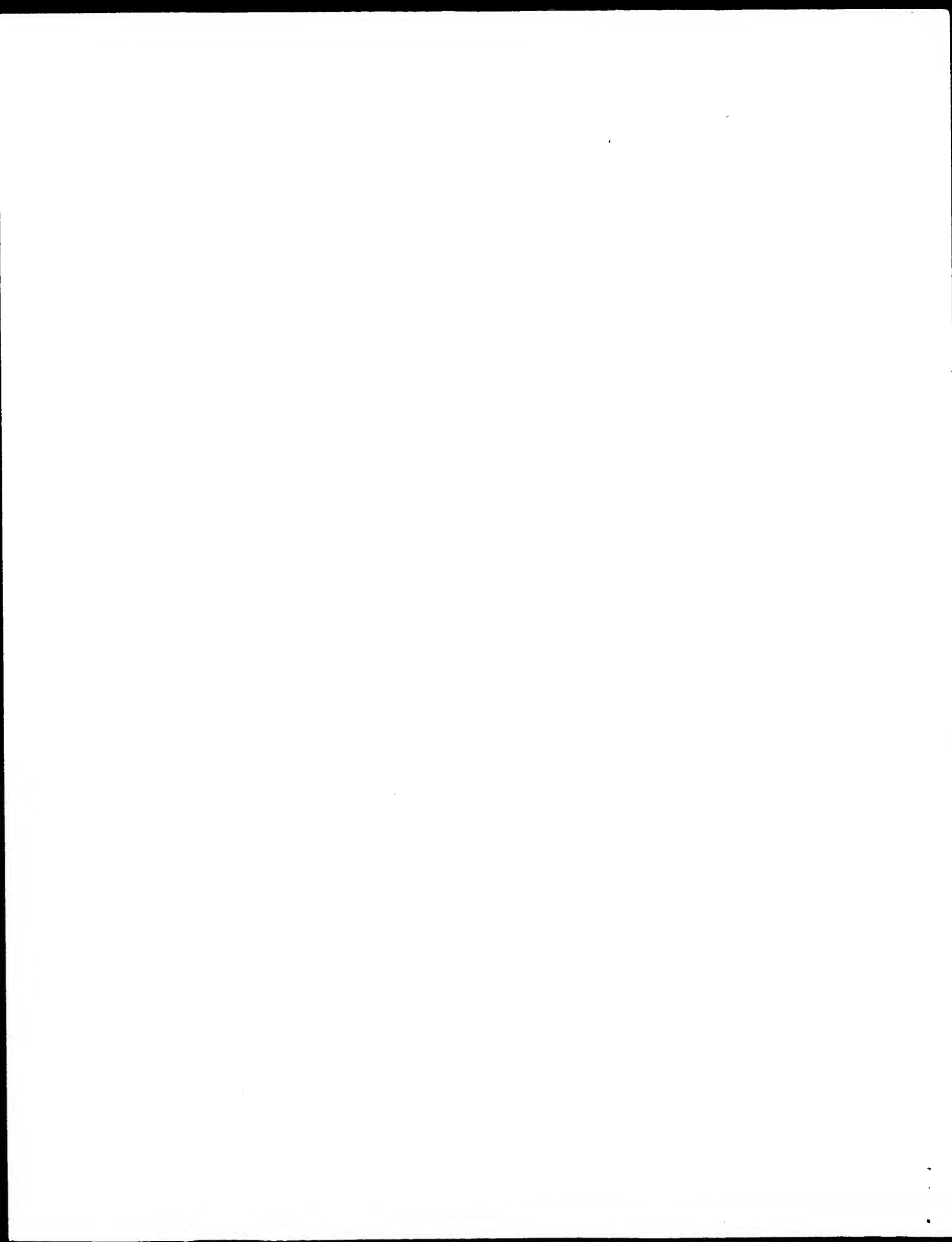
Search information block:

Query: US-09-674-779-4
Query length: 60
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 16.530000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig zScore EScore Len ! Documentation ..



OM of: US-09-674-779-4 to: SPTREMBL_19:* out_format : pfs

Date: Jul 30, 2002 4:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

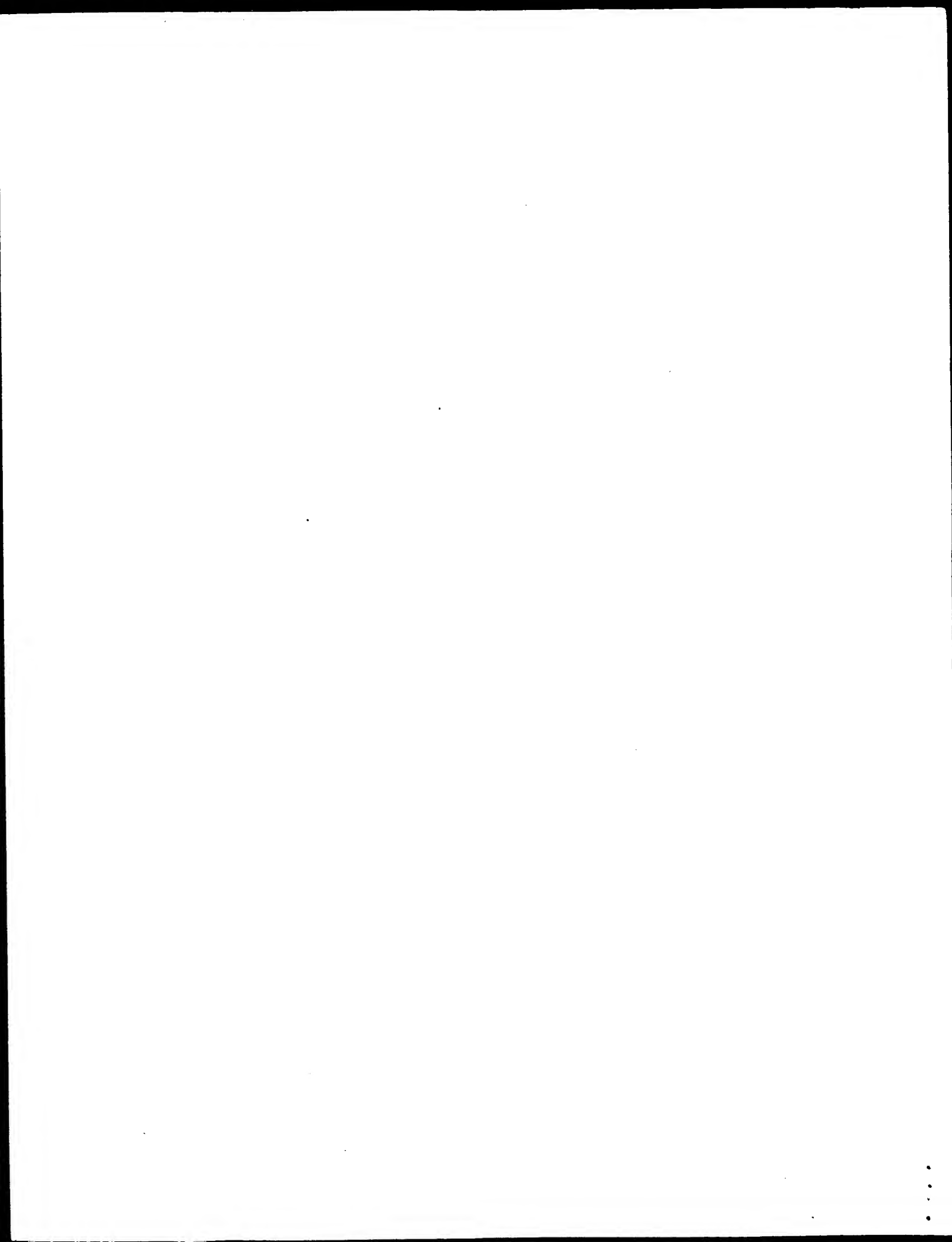
```
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09674779/runat_30072002_151837_8596/app_query.fasta_1.116
-DB=SPTREMBL_19 -CFMT=fastan -SUFFIX=oligo.rspt -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGN1_1_205 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: SPTREMBL_19:*
Database sequences: 56222
Database length: 172994929
Search time (sec): 47.120000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:	Strd Orig	ZScore	EScore	Len	! Documentation	..
Sequence						



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:07:07 ; Search time 29.83 Seconds
(without alignments)
930.890 Million cell updates/sec

Title: US-09-674-779-2

Perfect score: 1355

Sequence: 1 MKNPNQYFITLISMLVAC.....GAQFSETNSTRHVLKPKNL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1355	100.0	250	22	AAB60645
2	89.5	6.6	1377	22	ABG12098
3	88	6.5	320	22	ABG12098
4	88	6.5	320	22	ABG12098
5	83.5	6.2	1218	22	ABG12098
6	83	6.1	932	22	ABG12098
7	83	6.1	932	22	ABG12098
8	82.5	6.1	842	22	ABG12098
9	82	6.1	308	21	AA174841
10	81	6.0	382	20	AA174841
11	81	6.0	1596	22	ABG12098

12	80	5.9	448	22	AAB73524	Human transferase
13	78	5.8	233	22	ABG12098	Novel human diago
14	78	5.8	304	21	AA174840	Neisseria meningit
15	78	5.8	388	21	AA174840	Neisseria meningit
16	78	5.8	578	16	AA175085	Arabidopsis thalia
17	78	5.8	578	16	AA175085	Oxalyl-CoA decarbo
18	77.5	5.7	307	21	AA175085	Human oxalyl-CoA d
19	77.5	5.7	342	21	AA175085	Arabidopsis thalia
20	77.5	5.7	358	21	AA175085	Arabidopsis thalia
21	77.5	5.7	2464	22	ABG12098	Arabidopsis thalia
22	77	5.7	308	21	AA175085	Drosophila melanog
23	77	5.7	330	22	AA175085	Neisseria gonorrhoe
24	77	5.7	330	22	AA175085	Amino acid sequenc
25	77	5.7	330	22	AA175085	Amino acid sequenc
26	77	5.7	368	20	AA175085	L. monocytogenes a
27	77	5.7	612	22	ABG12098	Drosophila melanog
28	77	5.7	935	21	AA175085	Protein encoded by
29	77	5.7	1783	22	ABG12098	Drosophila melanog
30	76.5	5.6	867	22	ABG12098	Novel human diago
31	76.5	5.6	884	22	AA175085	Human polypeptide
32	76.5	5.6	899	22	AA175085	Human protein SEQ
33	76.5	5.6	903	22	AA175085	Human polypeptide
34	76.5	5.6	961	22	AA175085	Human polypeptide
35	76.5	5.6	1798	21	AA175085	Human HSGT1 protei
36	76	5.6	352	21	AA175085	Arabidopsis thalia
37	76	5.6	1139	22	ABG12098	Drosophila melanog
38	75.5	5.6	300	22	ABG12098	Drosophila melanog
39	75.5	5.6	3722	12	AA175085	Cephalosporin anti
40	75.5	5.6	3972	21	AA175085	S. avermitilis ave
41	75.5	5.6	3972	22	AA175085	Streptomyces averm
42	75.5	5.6	3972	22	AA175085	Streptomyces averm
43	75	5.5	306	20	AA175085	Topoisomerase II b
44	75	5.5	337	21	AA175085	Human cancer assoc
45	75	5.5	362	21	AA175085	Human CD40 recepto

ALIGNMENTS

RESULT 1

ID AAB60645 standard; Protein; 250 AA.

AC AAB60645;

DT 04-MAY-2001 (first entry)

DE Moraxella catarrhalis strain ATCC43617 BASB120 protein.

KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory.

CS Moraxella catarrhalis.

XX WO200109335-A2.

PN 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EF07361.

XX 03-AUG-1999; 99GB-0018281.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonard J;

DR WPI; 2001-159872/16.

XX N-PSDB; AAF59797.

XX New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT strain American Type Culture Collection 43617, for use as therapeutic

[illegible]

Db 230 svppmeyrhsmatgaptssmnlhmsspttkamtssviyhp-----pamshhts 281

QY 173 NSAIWIW 180

Db 282 lssstppwm 289

RESULT 9

AAV74841

ID AAV74841 standard; Protein; 308 AA.

XX

AC AAV74841;

XX

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 280 protein sequence SEQ ID NO:1156.

XX

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

KW antibacterial; gene therapy.

XX

OS Neisseria meningitidis.

XX

XX WO9957280-A2.

XX

PD 11-NOV-1999.

XX

XX 30-APR-1999; 99WO-US09346.

PF

XX 01-MAY-1998; 98US-0083758.

PR

XX 31-JUL-1998; 98US-0094869.

PR

XX 02-SEP-1998; 98US-0098994.

PR

XX 02-SEP-1998; 98US-0099062.

PR

XX 09-OCT-1998; 98US-0103749.

PR

XX 09-OCT-1998; 98US-0103794.

PR

XX 09-OCT-1998; 98US-0103796.

PR

XX 25-FEB-1999; 99US-0121528.

XX

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX

XX WPI: 2000-062150/05.

DR

DR N-PSDB; AA253603.

XX

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

XX

XX Claim 2; Page 638; 1453pp; English.

XX

XX AA253015 to AA254536, AA254577 to AA254615, and AA254615 to AA254941

CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the

CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

XX

XX Sequence 308 AA;

SQ

Query Match 6.1%; Score 82; DB 21; Length 308;

Best Local Similarity 19.9%; Pred. No. 1.6;

Matches 41; Conservative 25; Mismatches 68; Indels 72; Gaps 9;

QY 30 VSIKTPSVLITKIDGHHHTHEHDESVHVGLOAHFETWLMHH-----74

Db 104 iqplk-----aeegghhhdhdhdhdhdeg-----hhhdgcydphwvndpv 146

QY 75 ---ATKQEVVRY-----QAYLQSRRLGNYLPPMSQL-----LTTARSWQACGH 113

Db 147 lmsayacnvaelialkadpegkvyvqgrlgnymqlkklhsdaqaafnavaakrkvltgh 206

QY 114 EPYQ-----LPPEHLWGOIVPTLH-----LVQDLKSRGILPANTO-----IRSVYR- 154

Db 207 dafsymgkryhiefiapqgvssaeapsakvaaiairqikregikavftenikdtrmvdri 266

QY 155 NPELNOCAGGAAMSKHLTNSAIDIW 180

Db 267 aketgvnsvgklysdalgnapadyi 292

RESULT 10

AAV37545

ID AAV37545 standard; Protein; 382 AA.

XX

AC AAV37545;

XX

DT 07-OCT-1999 (first entry)

DE

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;

KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

XX Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX

XX WO9928475-A2.

PN

XX 10-JUN-1999.

PD

XX 27-NOV-1998; 98WO-IB01939.

PF

XX 04-NOV-1998; 98US-0107077.

PR

XX 28-NOV-1997; 97FR-0015041.

PR

XX 17-DEC-1997; 97FR-0016034.

XX

XX (GEST) GENSET.

PA

XX Griffais R;

PI

XX WPI: 1999-371125/31.

DR

XX Genome sequence of Chlamydia trachomatis

PT

XX Disclosure; Page 1207-1209; 1755pp; English.

PS

XX AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perihhepatitis, Bartholinitis; pneumopathy in breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX

XX Sequence 382 AA;

SQ

Query Match 6.0%; Score 81; DB 20; Length 382;

Best Local Similarity 18.9%; Pred. No. 2.9;

Matches 50; Conservative 40; Mismatches 90; Indels 84; Gaps 11;

QY 17 LVACAPITNPQVS-----PIKPSVLITKDKIGDHHTHE 52
 Db 132 lissktptntnyisafeslytdvqsvskrletlqkdpattptfsvfsqqlfhdplhp 191
 QY 53 HDESV-----SHVGLQAHFETWLOMHHTKQEVVR 82
 Db 192 ndqgvppllslplknhlqgtihgvcsgvyflfsmhpg eafkkt--qkthgfpkdafe 248
 QY 83 YOAYLQSLGNLPPMSQLL--TTARSQACGHE-----PYQLPPEHLWGQIVPTLHLYQ 135
 Db 249 leslyesylsaligashlplkeatlirvtrphdiepllpfssesyl----irahlyd 304
 QY 136 DLKSGRI--LPANTQIRSVYRNPELNQACGAAMSKHLTNSAID-----IWVDPDL---EI 185
 Db 305 llqnvqvpplsagaweylkktdesasci-----kkildseedltsfswfhnrdteel 358
 QY 186 KSQLYELQNLCOYWLEHGENQN 209
 Db 359 cseisndvlsqllsifinqdsqgn 382

RESULT 11
 ABB70845
 ID ABB70845 standard; Protein; 1596 AA.
 AC ABB70845;
 XX
 XX
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 39327.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL14948.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 39327; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB857737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1596 AA;

Query Match 6.0%; Score 81; DB 22; Length 1596;
 Best Local Similarity 21.2%; Pred. No. 24; Mismatches 109; Indels 68; Gaps 13;
 Matches 58; Conservative 38; Mismatches 109; Indels 68; Gaps 13;
 QY 11 TLISSMLVACSAPIPTNPQVSPKPT--PSVLITKDKIGDHH-----THEHD 54
 Db 965 tlmhssmlqcss--tsttlatittstmpdmqppqqlvhygavlhplhqqhgehqqrq 1023
 QY 55 ESVSHVGLQAHFETWLOMHHTKQEVVRYOAYLQSLR-----GNVLPMSQLLTTA 105
 Db 1024 ea-----dhqqqrrelhldqgqqqgqaladslphssssstssppmpmplttti 1076
 QY 106 RSWQACGHEPYQLPPEHLWG--QIVPTLHL-----YDLKSGRIPLPANTQIRSVVR 154
 Db 1077 tapqll---plpppphitstmpmptmhmpimppppqcyqqlqp--ldptmsyhtilgs 1131
 QY 155 NPELNQACGAAMSKHLTNSAIDIWVDPLEIKSQALYELQNLCOYWLEHGENQNFG--- 211
 Db 1132 gpeahgttagggysnqittsdqgi---lqlmptslfapypaplspsysvaagrsqegdlp 1187
 QY 212 ----LGLYATGAIHLDTGFRKWKGAQFSETNSI 240
 Db 1188 pvhtl-----ttalhahqgg---gqgeaqtpti 1212

RESULT 12
 AAB73524
 ID AAB73524 standard; Protein; 448 AA.
 XX
 XX AC AAB73524;
 XX
 XX DT 31-JUL-2001 (first entry)
 XX
 XX DE Human transferase HTFS-31, SEQ ID NO:31.
 XX
 KW Human transferase; HTFS; agonist; antagonist; cellular signalling;
 KW proliferation; cell proliferative disorder; immune disorder;
 KW atherosclerosis; hepatitis; psoriasis; cancer; tumour;
 KW inflammation; AIDS; Addison's disease; allergy; asthma; anaemia;
 KW cirrhosis; Crohn's disease; atopic dermatitis; diabetes mellitus;
 KW multiple sclerosis; rheumatoid arthritis; pancreatitis;
 KW systemic lupus erythematosus; thrombocytopenia; ulcerative colitis;
 KW haemodialysis; extracorporeal circulation; trauma; transgenic animal;
 XX gene therapy; drug screening.
 OS Homo sapiens.
 XX
 PN WO200132888-A2.
 XX
 PD 10-MAY-2001.
 XX
 XX 02-NOV-2000; 2000WO-US30485.
 XX
 PR 04-NOV-1999; 99US-0163595.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Tang YT, Yue H, Hillman JL, Lal P, Bandman O, Patterson C;
 PI Shih LL, Azimzal Y, Lu DAM, Baughn MR;
 XX
 DR WPI; 2001-328796/34.
 XX N-PSDB; AAB73524.
 XX
 XX Human transferase polypeptides and polynucleotides useful for
 XX diagnosis, prevention and treatment of cell proliferative and immune
 XX system disorders and for identifying agonists and antagonists -
 PS Claim 1; Page 126-127; 157pp; English.
 XX
 CC Sequences AAB73494-AAB73535 represent novel human transferase proteins
 CC HTFS-1 to HTFS-42, and sequences AAB73801-AAB73842 represent cDNAs
 CC encoding them. The proteins play important roles in the regulation of


```

XX OS  Neisseria meningitidis.
XX PN  W09957280-A2.
XX PD  11-NOV-1999.
XX PF  30-APR-1999; 99WO-US09346.
XX PR  01-MAY-1998; 98US-0083758.
XX PR  31-JUL-1998; 98US-0094869.
XX PR  02-SEP-1998; 98US-0098994.
XX PR  02-SEP-1998; 98US-0099062.
XX PR  09-OCT-1998; 98US-0103749.
XX PR  09-OCT-1998; 98US-0103794.
XX PR  09-OCT-1998; 98US-0103796.
XX PR  25-FEB-1999; 99US-0121528.
XX PA  (CHIR ) CHIRON CORP.
XX PA  (GENO-) INST GENOMIC RES.
XX PI  Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX PI  Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI  Tettelin H, Venter JC;
XX DR  WPI; 2000-062150/05.
XX DR  N-PSDB; AA253602.
XX PT  Novel Neisserial polypeptides predicted to be useful antigens for
XX PT  vaccines and diagnostics -
XX PS  Claim 2; Page 636-637; 1453pp; English.
XX CC  AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA25941
XX CC  represent novel Neisseria meningitis and N. gonorrhoea polynucleotides
XX CC  and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX CC  PCR primers used in the exemplification of the present invention. The
XX CC  polypeptides, the polynucleotides, antibodies and compositions of
XX CC  the invention can be used as vaccines, as diagnostic reagents, and as
XX CC  immunogenic compositions. The polypeptides can be used in the
XX CC  manufacture of medicaments for treating or preventing infection due to
XX CC  Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX CC  presence of Neisseria bacteria, or to raise antibodies. They may also
XX CC  be used to screen for agonists or antagonists, which may themselves
XX CC  have use as antibacterial agents. The polynucleotides of the invention
XX CC  may also be used in gene therapy protocols.
XX SQ  Sequence 304 AA;

Query Match      5.8%; Score 78; DB 21; Length 304;
Best Local Similarity 21.1%; Pred. No. 4.5;
Matches 41; Conservative 27; Mismatches 74; Indels 52; Gaps 10;

Qy 30 VSPKTPSVLTKDKIGDHTHEHDSVSHVGLQAHFE----TWLQ--MHATKQEVVRY 83
Db 104 iqlk-----aeegghdhhdhdheghh---hdhgeyqphwvndpvlmsayaqnvaka 154
Qy 84 -----QAYLQSRGLNPLPPMSOL-----LTTARSHWQACGHEPYQ----- 117
Db 155 likadpegkvyqqrlngymqikllhsdaqaafnavpaakrkvltghdaafsymgkryhi 214
Qy 118 --LPPEHLWGQIVPTLH----LYQDLKSRGILPANTO-----IRSVYR-NPELNQACAGAA 166
Db 215 efiapqgvsseaeapsakvaairkqregikavftenikdrtmvdriaketsvnsvgl 274
Qy 167 MSKHLTNSADIWV 180
Db 275 ysdalgnapadyi 288

RESULT 15
AAG41671

```

```

ID XX  AAG41671 standard; Protein; 388 AA.
XX AC  AAG41671;
XX XX  18-OCT-2000 (first entry)
XX DT  Arabidopsis thaliana protein fragment SEQ ID NO: 51877.
XX DE  Protein identification; signal transduction pathway; metabolic pathway;
XX XX  hybridisation assay; genetic mapping; gene expression control; promoter;
XX XX  termination sequence.
XX OS  Arabidopsis thaliana.
XX PN  EP1033405-A2.
XX PD  06-SEP-2000.
XX PF  25-FEB-2000; 2000EP-0301439.
XX PR  25-FEB-1999; 99US-0121825.
XX PR  05-MAR-1999; 99US-0123180.
XX PR  09-MAR-1999; 99US-0123548.
XX PR  23-MAR-1999; 99US-0125788.
XX PR  25-MAR-1999; 99US-0126264.
XX PR  23-MAR-1999; 99US-0126785.
XX PR  01-APR-1999; 99US-0127462.
XX PR  06-APR-1999; 99US-0128234.
XX PR  08-APR-1999; 99US-0128714.
XX PR  16-APR-1999; 99US-0129845.
XX PR  19-APR-1999; 99US-0130077.
XX PR  21-APR-1999; 99US-0130449.
XX PR  23-APR-1999; 99US-0130510.
XX PR  28-APR-1999; 99US-0130891.
XX PR  30-APR-1999; 99US-0131449.
XX PR  30-APR-1999; 99US-0132048.
XX PR  04-MAY-1999; 99US-0132407.
XX PR  05-MAY-1999; 99US-0132484.
XX PR  06-MAY-1999; 99US-0132485.
XX PR  06-MAY-1999; 99US-0132486.
XX PR  07-MAY-1999; 99US-0132487.
XX PR  11-MAY-1999; 99US-0132863.
XX PR  14-MAY-1999; 99US-0134256.
XX PR  14-MAY-1999; 99US-0134218.
XX PR  14-MAY-1999; 99US-0134219.
XX PR  14-MAY-1999; 99US-0134221.
XX PR  18-MAY-1999; 99US-0134370.
XX PR  19-MAY-1999; 99US-0134768.
XX PR  20-MAY-1999; 99US-0134941.
XX PR  21-MAY-1999; 99US-0135124.
XX PR  24-MAY-1999; 99US-0135353.
XX PR  25-MAY-1999; 99US-0135629.
XX PR  27-MAY-1999; 99US-0136021.
XX PR  28-MAY-1999; 99US-0136392.
XX PR  01-JUN-1999; 99US-0136782.
XX PR  03-JUN-1999; 99US-0137222.
XX PR  04-JUN-1999; 99US-0137528.
XX PR  07-JUN-1999; 99US-0137502.
XX PR  08-JUN-1999; 99US-0137724.
XX PR  10-JUN-1999; 99US-0138094.
XX PR  10-JUN-1999; 99US-0138540.
XX PR  14-JUN-1999; 99US-0138847.
XX PR  16-JUN-1999; 99US-0139119.
XX PR  16-JUN-1999; 99US-0139452.
XX PR  16-JUN-1999; 99US-0139453.
XX PR  17-JUN-1999; 99US-0139492.
XX PR  18-JUN-1999; 99US-0139454.
XX PR  18-JUN-1999; 99US-0139455.
XX PR  18-JUN-1999; 99US-0139456.
XX PR  18-JUN-1999; 99US-0139457.
XX PR  18-JUN-1999; 99US-0139458.
XX PR  18-JUN-1999; 99US-0139459.
XX PR  18-JUN-1999; 99US-0139460.

```

PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144086.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144335.	PR	13-OCT-1999;	99US-0159295.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144684.	PR	14-OCT-1999;	99US-0159331.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0145088.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160768.
PR	23-JUL-1999;	99US-0145145.	PR	21-OCT-1999;	99US-0160770.
PR	23-JUL-1999;	99US-0145218.	PR	21-OCT-1999;	99US-0160814.
PR	26-JUL-1999;	99US-0145276.	PR	21-OCT-1999;	99US-0160815.
PR	27-JUL-1999;	99US-0145913.	PR	22-OCT-1999;	99US-0160980.
PR	27-JUL-1999;	99US-0145918.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145919.	PR	22-OCT-1999;	99US-0160989.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161404.
PR	02-AUG-1999;	99US-0146386.	PR	25-OCT-1999;	99US-0161405.
PR	02-AUG-1999;	99US-0146388.	PR	25-OCT-1999;	99US-0161406.
PR	03-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161359.
PR	04-AUG-1999;	99US-0147038.	PR	26-OCT-1999;	99US-0161360.
PR	04-AUG-1999;	99US-0147204.	PR	26-OCT-1999;	99US-0161361.
PR	05-AUG-1999;	99US-0147192.	PR	28-OCT-1999;	99US-0161920.
PR	05-AUG-1999;	99US-0147260.	PR	28-OCT-1999;	99US-0161992.
PR	06-AUG-1999;	99US-0147303.	PR	28-OCT-1999;	99US-0161993.
PR	06-AUG-1999;	99US-0147416.	PR	29-OCT-1999;	99US-0162142.
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			

Query Match

Best Local Similarity

Matches

5.8%;

21.8%;

65;

Score 78;

DB 21;

Length 388;

Pred. No. 6.5;

Mismatches 32;

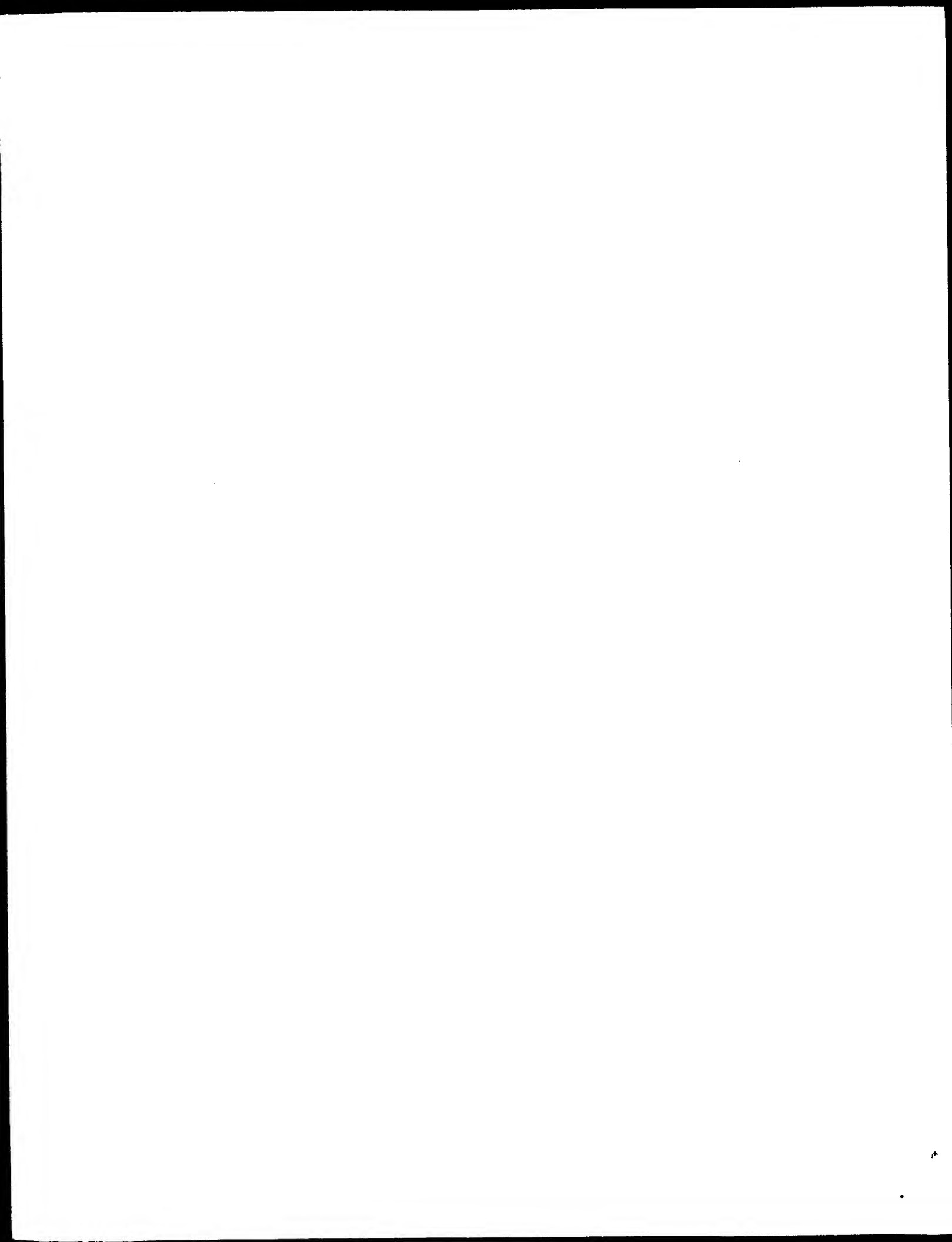
Indels 114;

Gaps 16;

QY	31	SPIKTP-----SVLITKDKTIGDHTHEHDES-----VSHVGLQAHFETWLQMH-HAT	76
Db	13	sslkkppllivilivcitvsvllitymvpqhnnssksacaglsrgcqaalsgwlpvhrkf	72
QY	77	KQEVRYQAYLQSLRGLNVLPPMSOLLTTARSWQACGH-EPYOLPPEHLWGOIVPTLHYQ	135
Db	73	tdeevaarvvikdlil-rilpp-----alcakskiafmlftpgtlpfeklwdk-----ffq	120
QY	136	DLKSR-GILPANTQIRSVYRNPELNQCAGGAMSKHLNSAIDIWVPDLKRSQ-----	188
Db	121	gcqgrfsiyihpsrlrpvh-----isrhfs-----dreihsdhvtwgr	158
QY	189	-ALYELONRLOCOYWLHEHGENQNF-----GLGLY	215

Db 159 ismvdarrillanaledpdnqhfvllsesciplhtfdytyrylmhanvsvfidsfedlgph 218
QY 216 ATGAIHLD-----TQGRK---W-----GAQFSETNSICRHVLPKNK 249
Db 219 gtgr-hmdhmlpeiprgdfrkgacqwtmkrqhavivmadgllyskfreycrpgveank 275

Search completed: July 30, 2002, 16:15:08
Job time: 481 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:07:52 ; Search time 13.04 Seconds
(without alignments)
468.282 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 1355
Sequence: 1 MNFNQYFTITLISMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	78	5.8	578	1	US-08-458-120-2
2	78	5.8	578	2	US-08-867-970-2
3	78	5.8	578	4	US-09-326-217-2
4	77	5.7	368	3	US-08-972-902-2
5	75	5.5	755	3	US-09-071-101-2
6	75	5.5	755	3	US-09-369-618-2
7	75	5.5	755	3	US-09-369-617-2
8	74	5.5	462	4	US-09-165-241-1
9	73	5.4	420	2	US-08-588-2588-40
10	73	5.4	420	5	PCT-US96-08285-40
11	73	5.4	1141	1	US-08-363-300-2
12	71.5	5.3	140	4	US-09-439-313-477
13	71	5.2	1091	3	US-08-986-485-5
14	70.5	5.2	305	1	US-08-844-085-6
15	70	5.2	437	4	US-09-353-332-2
16	70	5.2	512	4	US-09-149-922-2
17	69	5.1	410	6	5177197-1
18	69	5.1	1394	6	5177197-30
19	68.5	5.1	436	3	US-08-660-347-2
20	68.5	5.1	473	2	US-08-700-670A-23
21	68.5	5.1	883	2	US-08-596-366-2
22	68.5	5.1	883	2	US-08-967-104-2
23	68	5.0	111	3	US-08-875-811-22
24	68	5.0	226	4	US-08-858-207A-417
25	68	5.0	473	4	US-09-286-691-15
26	68	5.0	473	4	US-09-687-147-15
27	68	5.0	657	3	US-08-893-852A-3

28	68	5.0	657	4	US-08-821-818-3	Sequence 3, Appli
29	68	5.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
30	67.5	5.0	622	2	US-08-459-146-2	Sequence 2, Appli
31	67.5	5.0	622	2	US-08-459-065-2	Sequence 2, Appli
32	67.5	5.0	712	2	US-08-468-576B-17	Sequence 17, Appli
33	67.5	5.0	712	2	US-08-468-579B-17	Sequence 17, Appli
34	67.5	5.0	712	3	US-08-468-577B-17	Sequence 17, Appli
35	67.5	5.0	820	4	US-09-173-914-2	Sequence 3, Appli
36	67.5	5.0	1245	2	US-08-801-263A-3	Sequence 3, Appli
37	67.5	5.0	1245	3	US-08-801-263A-6	Sequence 6, Appli
38	67.5	5.0	1245	3	US-09-102-248-3	Sequence 3, Appli
39	67.5	5.0	1245	3	US-09-102-248-6	Sequence 3, Appli
40	67.5	5.0	2485	5	PCT-US94-00198-1	Sequence 1, Appli
41	67.5	5.0	2485	5	PCT-US94-00198-2	Sequence 2, Appli
42	67.5	5.0	2818	1	US-08-510-284-1	Sequence 1, Appli
43	67.5	5.0	2818	1	US-08-411-389-2	Sequence 2, Appli
44	67.5	5.0	2818	2	US-08-449-933-2	Sequence 2, Appli
45	67.5	5.0	2818	4	US-07-966-049A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-458-120-2
; Sequence 2, Application US/08458120
; Patent No. 5635616
; GENERAL INFORMATION:
; APPLICANT: OLSEN, ET AL.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,120
; FILING DATE: SUBMITTED HEREWITH
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05561
; FILING DATE: 18 MAY 94
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 3258000-312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-458-120-2

Query Match 5.8%; Score 78; DB 1; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 24; Mismatches 71; Indels 58; Gaps 11;

QY 82 RYQ---AYLQ-----SRILNYLPPMSQLL-----TTARSWQACGHEPVLPPHLWGOIV 128

Db 294 RYQDVKFIQVDICAEELGNNVKKPAVTLGNHATVKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176
Db 352 -TLREKMKSEAAASKELASKSLPMNYTVFVHVQEQLPDRCFVWSEGA-----NTM 402
QY 177 DIWVDPLEIKSQAELYELQNLRCQWLEHGENQFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNYLPRHLDAGTGTGTMGVGLGFAIAAAVAKDRSPGOWIICV 452
QY 225 QGFRKKGAFSETNSICRHVLP 246
Db 453 EGDSAFGSGMEVETICRYNLP 474

RESULT 2
US-08-867-970-2
; Sequence 2, Application US/08867970
; Patent No. 5945273
; GENERAL INFORMATION:
; APPLICANT: Olsen, Henrik S.
; APPLICANT: Coleman, Timothy
; APPLICANT: Adams, Mark D.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,970
; FILING DATE: JUN-3-97
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/458,120
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/05561
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF119D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-867-970-2

Query Match 5.8%; Score 78; DB 2; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 24; Mismatches 71; Indels 58; Gaps 11;
QY 82 RYQ---AYLQ-----SRIGNVLPPMSQL-----TTARSWQACGHEPYQLPPEHLGQIV 128
Db 294 RYQDVKFIQVDICAEELGNNVKKPAVTLGNHATVKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176

Db 352 -TLREKMKSEAAASKELASKSLPMNYTVFVHVQEQLPDRCFVWSEGA-----NTM 402
QY 177 DIWVDPLEIKSQAELYELQNLRCQWLEHGENQFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNYLPRHLDAGTGTGTMGVGLGFAIAAAVAKDRSPGOWIICV 452
QY 225 QGFRKKGAFSETNSICRHVLP 246
Db 453 EGDSAFGSGMEVETICRYNLP 474

RESULT 3
US-09-326-217-2
; Sequence 2, Application US/09326217
; Patent No. 6200796
; GENERAL INFORMATION:
; APPLICANT: Olsen, Henrik S.
; APPLICANT: Coleman, Timothy
; APPLICANT: Adams, Mark D.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/326,217
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,970
; FILING DATE: JUN-03-97
; APPLICATION NUMBER: US 08/458,120
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/05561
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF119D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-326-217-2

Query Match 5.8%; Score 78; DB 4; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 24; Mismatches 71; Indels 58; Gaps 11;
QY 82 RYQ---AYLQ-----SRIGNVLPPMSQL-----TTARSWQACGHEPYQLPPEHLGQIV 128
Db 294 RYQDVKFIQVDICAEELGNNVKKPAVTLGNHATVKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176
Db 352 -TLREKMKSEAAASKELASKSLPMNYTVFVHVQEQLPDRCFVWSEGA-----NTM 402

QY 177 DIWVPDIKSQLYELQNLCOYWLHGENQNFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNYLPHRLDAGTFTGTMGVGLGFAIAAAVAVAKDRSPGOWIICV 452
QY 225 QGFRKWAQFSETNSTRVLP 246
Db 453 EGDSAFGFGMEVETICRYNLP 474

RESULT 4
US-08-972-902-2
; Sequence 2, Application US/08972902
; Patent No. 6099848
; GENERAL INFORMATION:
; APPLICANT: Frankel, Fred R.
; APPLICANT: Portnoy, Daniel A.
; TITLE OF INVENTION: BACTERIAL VACCINE VECTOR AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: FL
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972.902
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle-Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-972-902-2

Query Match 5.7%; Score 77; DB 3; Length 368;
Best Local Similarity 21.3%; Pred. No. 2.3;
Matches 46; Conservative 34; Mismatches 98; Indels 38; Gaps 9;

QY 50 THEHDESV---SHVQLQAHFETLQMHATKQEVVRYQAYLQSRGLGNYLPPMSQLTT-- 104
Db 90 TRKEDANLAANKHISLTVFEDWLE--NLTLATLRLHLKVDSCMGL-----GIRTEE 142
QY 105 ARSQWACGHEPYQLPPEHLWGQIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNOCAGG 164
Db 143 ARREATSTNDHQLQLEGIVTHFAATQLETSVFEQOLAKFOTILSLKKRPTVYHTANS 202
QY 165 AA---MSKHLTNSAIDIW-----PDLEIKSQALYELQNLCOYW-----LEHGENON 209
Db 203 AASLLQIQIGDFAIRFGISMYGLTPSTEIKTSLPFELKPALALYTEMVHVVKELAPGDSVS 262
QY 210 FGLGLYATGAIHLDTOGFRKWA--QFSETNSTRH 243

Db 263 YGATYATATE-----REWVATLPIGYADGLIRH 289

RESULT 5
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,101
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E.
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-071-101-2

Query Match 5.5%; Score 75; DB 3; Length 755;
Best Local Similarity 22.0%; Pred. No. 12;
Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;

QY 14 SSMVLVACAPITPTNOVSPIKTPSVLITKDKGDHH--THEHDESVSHVGLQA----- 64
Db 340 STLTITTSYSGVATDQI-----VTTDL---HHGCTDQHTGTSASAPLAAGMIALAL 387
QY 65 ---HFEETLQMHATKQEVVRYQ--AYLOS-----RLGNYLPPMSQLTTA 105
Db 388 EANPFLUTRDMQHL-----VVRASKFAHLQAEADWRNTRNGVRQVSHHYGYGLLDAGLLVDTA 443
QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYQDLKSR--GILPANTQIRSVYRNPELNOCAG 163
Db 444 RTW-----LPTQ-----PQKCAVRVQSRPTLPL-----IYIRENVVSACAG 481
QY 164 GRAMSKHLTNSAIDIWDPDLEIKSQALYELQNLCOYWLHGE-----NQNFGLG-----LYA 216
Db 482 -----LHNSI-----RSLEHVQAQLTSLYSRRGDLTSLTSPMGTRSLVA 522
QY 217 TGAIHLDTOGFRK 230
Db 523 IRPLDVSTEGYNNW 536

RESULT 6
US-09-369-618-2
; Sequence 2, Application US/09369618
; Patent No. 6100041
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D2
; CURRENT APPLICATION NUMBER: US/09/369,618
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-369-618-2

Query Match 5.5%; Score 75; DB 3; Length 755;
Best Local Similarity 22.0%; Pred. No. 12;
Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;
QY 14 SSMVLACSAPIPTNPQVSPKTPSVLITKDKIGDHH--THEHDESVSHVGLQA-----64
Db 340 STLTITYSSGVATDPOI-----VTTDL---HHGCTDQHTGTSASAPLAAGMIALAL 387
QY 65 ---HFETWLOMHATKQEVVRQ--AYLOS-----RLGNVLPMPMSOLLTTA 105
Db 388 EAMPFLTWDRMQHL-----VVRASKPAHLQAEDWRTNGVRQVSHHYGGLDAGLLVDTA 443
QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYQDLKSR--GILPANTQIRSVYRNPELNOCAG 163
Db 444 RTW-----LPTQ-----PQRKCAVRVOSRPTPILPL-----IYIRENVSAACAG 481
QY 164 GAAMSKHLTNSAIDIWVPDLKTSQALYELQNLRCQYWLEHGE---NQNFGLG---LYA 216
Db 482 -----LHNSI-----RSLHVVQAQLTSLYSRRGDLSEISLTSPMGTRSTLVA 522
QY 217 TGAHLDTQGRKW 230
Db 523 IRPLDVSTEGYNNW 536

RESULT 7
US-09-369-617-2
; Sequence 2, Application US/09369617
; Patent No. 6127162
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D1
; CURRENT APPLICATION NUMBER: US/09/369,617
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-369-617-2

Query Match 5.5%; Score 75; DB 3; Length 755;
Best Local Similarity 22.0%; Pred. No. 12;
Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;
QY 14 SSMVLACSAPIPTNPQVSPKTPSVLITKDKIGDHH--THEHDESVSHVGLQA-----64
Db 340 STLTITYSSGVATDPOI-----VTTDL---HHGCTDQHTGTSASAPLAAGMIALAL 387
QY 65 ---HFETWLOMHATKQEVVRQ--AYLOS-----RLGNVLPMPMSOLLTTA 105
Db 388 EAMPFLTWDRMQHL-----VVRASKPAHLQAEDWRTNGVRQVSHHYGGLDAGLLVDTA 443
QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYQDLKSR--GILPANTQIRSVYRNPELNOCAG 163
Db 444 RTW-----LPTQ-----PQRKCAVRVOSRPTPILPL-----IYIRENVSAACAG 481
QY 164 GAAMSKHLTNSAIDIWVPDLKTSQALYELQNLRCQYWLEHGE---NQNFGLG---LYA 216
Db 482 -----LHNSI-----RSLHVVQAQLTSLYSRRGDLSEISLTSPMGTRSTLVA 522
QY 217 TGAHLDTQGRKW 230
Db 523 IRPLDVSTEGYNNW 536

RESULT 8
US-09-165-241-1
; Sequence 1, Application US/09165241
; Patent No. 6130077
; GENERAL INFORMATION:
; APPLICANT: Henry Yue
; APPLICANT: Karl J. Guegler
; APPLICANT: Neil C. Corley
; APPLICANT: Janice Au-Young
; TITLE OF INVENTION: HUMAN CYTOCHROME P450
; FILE REFERENCE: PF-0608 US
; CURRENT APPLICATION NUMBER: US/09/165,241
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 991729
US-09-165-241-1

Query Match 5.5%; Score 74; DB 4; Length 462;
Best Local Similarity 25.2%; Pred. No. 7.2;
Matches 35; Conservative 21; Mismatches 67; Indels 16; Gaps 7;
QY 113 HEPYQLPPEHLW--GOIVPTLHLYQDLK-----SRGILPANTQIRSVYRNPELNOCAGGA 165
Db 62 HERYG-PVVSFWFGRRLLVVSGLTVDVLKQHLNPKTLDPETMLKSLR-----YQSGGGS 116
QY 166 AMSKHLTNSAIDIWVPD--LEIKSQALYELQNLRCQYWLEHGENQNFGLGYATG-AIHL 223
Db 117 VSENHMRKKLYENGVTDSLKSNFALLKLSBELLDKWLSPETQHVPLSQHMLGFAMKSV 176
QY 224 TQGRKWAQFSETNSICR 242
Db 177 TQ--MVMGSTFDDQOEVR 193
RESULT 9
US-08-588-258B-40
; Sequence 40, Application US/08588258B
; Patent No. 5929207
; GENERAL INFORMATION:

QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYQDLKSRGILPANTOIRSVYRNPELNOCAGGA 165
Db 106 R-----HHHADTPP--LWCR-----LNYPAGGT 126
QY 166 AMSKHLTNSAIDWV 180
Db 127 AVA-----YSLSDWL 137
RESULT 13
US-08-986-485-5
; Sequence 5, Application US/08986485
; Patent No. 6046030
; GENERAL INFORMATION:
; APPLICANT: WU, SHUJIAN
; APPLICANT: SWEET, RAYMOND
; APPLICANT: TRUNEH, ALEMEGED
; TITLE OF INVENTION: A HUMAN LIG-1 HOMOLOG (HLIG-1)
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,485
; FILING DATE: 08-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/059,448
; FILING DATE: 22-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1091 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-986-485-5
Query Match 5.2%; Score 71; DB 3; Length 1091;
Best Local Similarity 20.6%; Pred. No. 62;
Matches 45; Conservative 35; Mismatches 88; Indels 50; Gaps 11;
QY 55 ESVSHVGL-----QAHFETWLQMHATKQEVRYQAYLQSRNLGNYLPP-----MSQLLT 103
Db 408 ESLEHLNLGENAIRSVQFADFAMKN-LKELYISSEFLCDQKWLPPWLMGRMLQAFV 466
QY 104 TARSQACGHEPYQLPPEHLWQ-----IVPTLHLYQDLKSRGILPANTOIRSVYRNPELN 159
Db 467 TA-----TCAH-----PELQSGSIFSVLPDSFVCDPFKQIITQPTTNAVVGKDIRF 516
QY 160 OCAGGAAMSKHLTNSAIDWVDPDLKQALYELQNLRCQYWLHGHENQNFGLGLYATGA 219
Db 517 TCSAASSSSPMTFA-----WKKDNEVLNA--DMEN-FAHVRADQGEVMEY-----TTI 563
QY 220 IHLDTQGF-----RKWGAQFSETNSICRHVLP 246

QY 13 ISSMLVACSAPIPTNPQVSPKIPSVLITKIGDHHHHEH---DESVHVG-----LQ 63
Db 555 LSHFPVCECAP-PEPAQSGPGVSQRKLMRYHSVSTETPHERKDFESKANHLGDTDGTVPK 613
QY 64 AHFETWLQ---MHATKQEVV---RYQAYLQSRNLGNYLPPMSQLLTAR----- 106
Db 614 TRRHSRQOIFLVAATPQKACDPSRYEDY--SELGE-LPPRSPLEPVCEGDFGQYRKK 670
QY 107 -----SWQACGHEP 115
Db 671 RGRHASFESCGKRP 685
RESULT 12
US-09-439-313-477
; Sequence 477, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 477
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-439-313-477
Query Match 5.3%; Score 71.5; DB 4; Length 140;
Best Local Similarity 21.5%; Pred. No. 2.1;
Matches 29; Conservative 14; Mismatches 43; Indels 49; Gaps 6;
QY 48 HHTHEHDESV--SHVGLQAHFETWLQMHATKQEVRYQAYLQSRNLGNYLPPMSQLLT 105
Db 50 HHTHEHTDTPYGHWHHTCHVTWTLHTITPPHTLP-----VDRTRHCHTDTQNTVTR 105

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AG3019
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-587 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AA44573.1; PID:g1742190; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu3763
 A:Map position: linear chromosome

Query Match 7.5%; Score 101; DB 2; Length 587;
 Best Local Similarity 22.5%; Pred. No. 0.64;
 Matches 38; Conservative 31; Mismatches 58; Indels 42; Gaps 9;
 QY 74 HATKQEVRYQAYLQSRIGNY----LPPMSQLLTARSQACGHEPYQLPPEHLWGQIVP 129
 Db HTREKAVITFK-----RNGKYDQKGLQELNRF--RDWRR--NQPTMDPR----- 69
 QY 130 TLHLYQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWPDLE 184
 Db 70 LFDLVWEVYRSGATDYINVVSFAFRSPETNGLLRTTKGVAEKSQHMLGKAMDFYIPG-- 127
 QY 185 IKSOALYELQNLCOYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
 Db 128 VKLATLEIGMQM-----QIGGVGYPTSGSPFVHMDVGGVRAW 166

RESULT 3
 B98265
 hypothetical protein AGR_L_2143 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 R:Accession: B98265
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 234, 2323-2328, 2001
 A:Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: B98265
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-634 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89644.1; PID:g15159543; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_2143
 A:Map position: linear chromosome

Query Match 7.5%; Score 101; DB 2; Length 634;
 Best Local Similarity 22.5%; Pred. No. 0.71; Mismatches 58; Indels 42; Gaps 9;
 Matches 38; Conservative 31; Mismatches 58; Indels 42; Gaps 9;
 QY 74 HATKQEVRYQAYLQSRIGNY----LPPMSQLLTARSQACGHEPYQLPPEHLWGQIVP 129
 Db 75 HTREKAVITFK-----RNGKYDQKGLQELNRF--RDWRR--NQPTMDPR----- 116
 QY 130 TLHLYQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWPDLE 184
 Db 117 LFDLVWEVYRSGATDYINVVSFAFRSPETNGLLRTTKGVAEKSQHMLGKAMDFYIPG-- 174
 QY 185 IKSOALYELQNLCOYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
 Db 175 VKLATLEIGMQM-----QIGGVGYPTSGSPFVHMDVGGVRAW 213

RESULT 4
 AH3510
 hypothetical membrane associated protein BMEII0010 [imported] - Brucella melitensis (str

C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AH3510
 R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
 ; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AH3510
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-637 <KUR>
 A:Cross-references: GB:AE008918; PIDN:AA53251.1; PID:g17984130; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEII0010
 A:Map position: II

Query Match 7.1%; Score 96.5; DB 2; Length 637;
 Best Local Similarity 22.9%; Pred. No. 1.8;
 Matches 40; Conservative 31; Mismatches 53; Indels 51; Gaps 12;
 QY 72 MHATKQEVRYQAYLQSRIGNYLP--MSQLLTARSQACGHEPYQLP--EHLWGQI 127
 Db 37 VHTGEKAEI---AFKKD--GRFLPDGLKRLNVLFLRDWRR--NEPTRMDPRFLDLI-- 85
 QY 128 VPTLHLQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVPD 182
 Db 86 ----QVQSTGSRVYI---TVVSAYRSPATNMLRSTRGVAKKSQHMLGRAMDYIPG 137
 QY 183 LEI-KSQAL---YELQNLCOYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
 Db 138 VPLAKLRAIGMRYQIG-----GVGYPRSGSPFVHMDVGNVRHW 176

RESULT 5
 AH0644
 Probable exported protein STY1255 [imported] - Salmonella enterica subsp. enterica se
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: This species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AH0644
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 , S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AH0644
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-263 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08339.1; PID:g16502384; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY1255

Query Match 6.9%; Score 93; DB 2; Length 263;
 Best Local Similarity 19.8%; Pred. No. 1.2;
 Matches 56; Conservative 37; Mismatches 92; Indels 98; Gaps 13;
 QY 9 ITTLSSMLVACSAPIPTNPQVSPKTP-SVLITKDKIGDHHTHEHDESVHVGLOAHFE 67
 Db 13 LLALATATMLA---LPAQANTWPLPPGSRVGENKF---HVVEDDGG----- 54
 QY 68 TWLQMHATKQEVRYQAYLQSRIG--NVLPPMSQLLTARSQACGHEPYQLPPEHLWG 125
 Db 55 ----SLEIAKKYNVGFLLQANPGIDPYVPRAGSVLT----- 89
 QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176

Db 90 --IPLQTLDPAPREGIVINIAELRLYYYPQKNTVTYPIGIGLQGGDTLTPTMTTIS 147
QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY----WLEHGENQ 208
Db 148 DKRANPTWTPTNIR--ARYKAQGIIDLPVAVPAGDPNMGHHAIRLAAYGGVYLLHGTNA 205
QY 209 NFGGLG-YATGAIHLDTQGRKWAQFSETNSICRHVLPKNKL 250
Db 206 GFGIGMRVSSGCIRL-----RGGDIETLFRQYTPGTRKV 238
RESULT 6
AD2215
two-component hybrid sensor and regulator all3275 [imported] - Anabaena sp. (strain PCC
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD2215
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2215
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-889 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074974.1; PID:g17132370; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3275

Query Match 6.6%; Score 90; DB 2; Length 889;
Best Local Similarity 20.2%; Pred. No. 9.9; Matches 45; Conservative 32; Mismatches 78; Indels 68; Gaps 10;
QY 15 SMLVACSAPITPNQVSPDKTPSVLITKDKI--GDHHTH-----EHDESYSVHVLQQA 64
Db 300 SLLVQCAVSLFVGIAITERSIOVELAVEKVRGEYQTRAELSEKLIQLNESLVAI---- 355
QY 65 HFETLQMHATKQEVRYQAYLQSRGLNVLPPMSQ--LLTARSQA--CGH----- 113
Db 356 --NDCLESREKDELQRCALSRSLNLSSTDAFIANRDWDITYCNHQAQIQL 413
QY 114 EPYOLPPEHLWGQIVPTLHLYODLKSRGILPANTQIRSVYRNPELNOQAGGAAMSKHLTN 173
Db 414 EPEDLICKNYWEQWTRT-----KGTDFEREYRRS----- 442
QY 174 SAIDIWVPLEIKSQALYELQNRLCOYWLE-HGENONFGLGLY 215
Db 443 -----LAENIPVHFELIYELWD-----MWLEVHAYPSRDLGLIF 476

RESULT 7
F64855
YcfS protein precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: F64855
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F64855
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <BLAT>
A:Cross-references: GB:AE000211; GB:U00096; NID:g1787345; PIDN:AAC74197.1; PID:g1787356;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ycfS

C:Superfamily: conserved hypothetical protein b0819
C:Keywords: periplasmic space
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-320/Product: ycfS protein #status predicted <MAT>

Query Match 6.5%; Score 88; DB 2; Length 320;
Best Local Similarity 20.8%; Pred. No. 4.1;
Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTLLISSMLVACSAPITPNQVSPDKTPSVLITKDKIGDHTHEDSHVSHVGLQAHPF 67
Db 10 WLTTFTFAAVALALPAKANTWPLP-PAGSRLVGENKF---HVENDGG----- 54
QY 68 TWLQMHATKQEVRYQAYLQSRGLG--NYLPPMSQLLTARSQACGHEPYQLPPEHLWG 125
Db 55 ---SLEATAKYNVGFALLQANPGVDVYPRAGSVLT----- 89
QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNOCA-----GGAAMSKHLTNSAI 176
Db 90 --IPLQTLDPAPREGIVINIAELRLYYYPGKNSVTYPIGIGLQGGDTLTPTMTTIS 147
QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY----WLEHGENQ 208
Db 148 DKRANPTWTPTNIR--ARYKAQGIELPAVVPAGLDNPMGHHAIRLAAYGGVYLLHGTNA 205
QY 209 NFGGLG-YATGAIHL 222
Db 206 DEFIGMRVSSGCIRL 220

RESULT 8
C90815
Hypothetical protein Ecs1491 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90815
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90815
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA034914.1; PID:g13360955; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs1491
C:Superfamily: conserved hypothetical protein b0819

Query Match 6.5%; Score 88; DB 2; Length 320;
Best Local Similarity 20.8%; Pred. No. 4.1;
Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTLLISSMLVACSAPITPNQVSPDKTPSVLITKDKIGDHTHEDSHVSHVGLQAHPF 67
Db 10 WLTTFTFAAVALALPAKANTWPLP-PAGSRLVGENKF---HVENDGG----- 54
QY 68 TWLQMHATKQEVRYQAYLQSRGLG--NYLPPMSQLLTARSQACGHEPYQLPPEHLWG 125
Db 55 ---SLEATAKYNVGFALLQANPGVDVYPRAGSVLT----- 89
QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNOCA-----GGAAMSKHLTNSAI 176
Db 90 --IPLQTLDPAPREGIVINIAELRLYYYPGKNSVTYPIGIGLQGGDTLTPTMTTIS 147
QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY----WLEHGENQ 208
Db 148 DKRANPTWTPTNIR--ARYKAQGIELPAVVPAGLDNPMGHHAIRLAAYGGVYLLHGTNA 205

QY 209 NFGLGL-YATGAIHL 222
 Db 206 DFGIGMRVSSGCIRL 220

RESULT 9
 G85674
 hypothetical protein ycf5 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85674
 R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 R.; Miller, L.; Grobeck, E.J.; Davis, A.; Dimmalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85674
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <STO>
 A:Cross-references: GB:AE005174; NID:g12514663; PIDN:AAG55859.1; GSPDB:GN00145; UWGP:Z17
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ycf5
 C:Superfamily: conserved hypothetical protein b0819

Query Match 6.5%; Score 88; DB 2; Length 320;
 Best Local Similarity 20.8%; Pred. No. 4.1;
 Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTISMLVACSAPIPTNPQVSPKTPSVLITDKIGDHTTHEDESVSHVGLQAHFE 67
 Db 10 WLTPFFAAVALALPAKANTWPLP-PAGSRLVGKPF---HVENDGG----- 54
 QY 68 TWLOMHATKQEVVRYOAVLYQSLRG--NYLPPMSQLLTITARSWQACGHPYOLPPEHLWG 125
 Db 55 ---SLEAIKKYVNGVFLALQANPGVDYVPRAGSVLT----- 89
 QY 126 QIVPTLHLVODLKSRLGILPANTQIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176
 Db 90 --IPLOTLDPDAPREGIVINIAELRLYYPPGKNSVTVPIGICQLGGDTLPPTWTVTS 147
 QY 177 D-----IWVPDLKISQALYELQN-----RLCOY---WLEHGENQ 208
 Db 148 DKRANPTWTPTANIR--ARYKAQGIELPAVVPAGPDNPMGHAIRLAAVGGVYLLHGTNA 205
 QY 209 NFGLGL-YATGAIHL 222
 Db 206 DFGIGMRVSSGCIRL 220

RESULT 10
 D71490
 probable exodeoxyribonuclease V, gamma - Chlamydia trachomatis (serotype D, strain UW3/C)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: D71490
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
 A:Reference number: A71570; MUID:99000809
 A:Accession: D71490
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1004 <ARN>
 A:Cross-references: GB:AE001334; GB:AE001273; NID:g3329078; PIDN:AAC68244.1; PID:g332908
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: recC

Query Match 6.5%; Score 88; DB 2; Length 1004;

Best Local Similarity 19.7%; Pred. No. 17;
 Matches 52; Conservative 40; Mismatches 88; Indels 84; Gaps 11;

QY 17 LVACSAPIPTNPQVS-----PIKTPSVLITDKIGDHTTHE 52
 Db 754 LLSKTPLTPTNYLSAFTESLYTDVQDSVSKRLETLOKDPATTPFSVTFSDQLFHDPLHP 813
 QY 53 HDESVS-----HVGLOAHFETWLOMHATKQEVVR 82
 Db 814 NDQOVSPLLLSLPLKNHIGTTHGVCVSGVYLFMSHPG-EAEKKT--OKTHGFPKDAFE 870
 QY 83 YOAYLQSRGLNYLPPMSQLL-TTARSWQACGHE-----PYQLPPEHLWGQIVPTLHLIQ 135
 Db 871 LESVLESYLSALLQASHLLPKKATILRVTPHDIEPLIPFPSPESYL-----IRAIHLYE 926
 QY 136 DLKSRGI-LPANTQIRSVYRNPELNCAGCAAMSKHLTNSAID-----IWVPDL---EI 185
 Db 927 LLQQAIPPLPSAOWEYIKTKTDSASQCI-----KKLLDSEEDPLTSSFWFHNRTTEEI 980
 QY 186 KSQALYELQNLRCQYWLHGENQN 209
 Db 981 CSRLSNDVLSQLSLFINQDSQON 1004

RESULT 11
 A11295
 phosphoribosylformylglycinamide synthetase I [imported] - Listeria monocytogenes (s
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: A11295
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
 Ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: A11295
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-739 <GIA>
 A:Cross-references: GB:NC_003210; PIDN:CAC99847.1; PID:g16411223; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: purQ
 C:Superfamily: phosphoribosylformylglycinamide synthase component II

Query Match 6.5%; Score 87.5; DB 2; Length 739;
 Best Local Similarity 19.5%; Pred. No. 13;
 Matches 48; Conservative 42; Mismatches 83; Indels 73; Gaps 13;

QY 40 ITKDKIGDHTTHEDESVSHVGLQAHFETWLOMHATKQEVVRYOAYLQSRGLNYLPPMS 99
 Db 352 VTDDKM--YKLIHGEVNVANVPDALAED-APVHKPSKEPTRYQAFQESEA--FVPMAD 406
 QY 100 QLLTTARSWQACGHPYOLPPEHLWGQIVPTLHLIQDLKSRGILP----- 144
 Db 407 DWVGV---WKELLAQPTIASKRRHIVEQ-----YDYQVTRDTAVVPGSDAAIVRVRGTEKA 458
 QY 145 ----ANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLEIKSQ 188
 Db 459 TAMTTDCNSRYLYLDPPEVGAIAVAEARNIVCSGGKPLA---ITDGLNFGNPE---KPE 512
 QY 189 ALYELQ---NRLCOYWLE-----HGE-----NONFGLGLYATGAI-----HLDTO 225
 Db 513 IFWEIEKAADGISACLELDTPVLSGNVSLYNETDGTGIYPTPIVGMVGLVEDLAHTTQ 572
 QY 226 GPRKMG 231
 Db 573 DFKNSG 578

QY 16 MLVACSAPIPTNPQVSPKTPSVLITKDKIGDHTTHEHDESVSHVQLQAHFETWLOMHA 75
:||| | | | | : | | : | | : | | :

hypothetical protein VC1750 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82162
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20408833
A:Accession: B82162
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <HEI>
A:Cross-references: GB:AE004252; GB:AE003852; NID:g9656263; PIDN:AAF94900.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1750
A:Map position: 1

Query Match		6.2%	Score 83.5;	DB 2;	Length 373;
Best Local Similarity		23.3%	Pred. No. 12;		
Matches	64;	Conservative	32;	Mismatches	88;
				Indels	91;
				Gaps	14;
QY	8	FITLLISSMLVACSA	-----IPTNP-----	QVSPIKTP--SV--LITKDKI	45
DB	3	FITLL-CFLLATLQPVQA	PTPIGLRCCAFGYNLHAQVAGIPVPFESVDNVIDVDAL	61	
QY	46	GDHHTHEHDESVS	-----HVGLOAHFFETWL-----	QMHH	74
DB	62	GGHYNQGDQSLSTSL	LLGSEHNGLIFTKRAFIDTAHVRTADFTYLFKLNLAQLGH	121	
QY	75	ATKQEVRYQAVYLOSRLGNV	-----LPPMSQLLTATRSWQACGHEPYQLPPEHLWGQIV	128	
DB	122	EAQ	---ITLPTELRSQIHWNQSVSLDPKQV---	RSAAQAAFIQFOLAQ---	WHEIA 172
QY	129	PTLHLY	-----QDLKSRGILPANTQIRSVYRNPELNOCAGGAAMSKHLTNS	174	
DB	173	QWFGLTAVVSGFHEQASAFSPEDLYS	-NMLGANLARDVLLANPDANKQEFKIFAHLEDE	231	
QY	175	AIDIWVPDLEIKSOALVEL	-----QNRICQVWL	202	
DB	232	LRKLKAPSSVTQOKIQOLEGIWWSQRRLPKWL	266		

Search completed: July 30, 2002, 16:15:54
Job time: 447 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:15:13 ; Search time 11.07 seconds
(without alignments)
874.425 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 1355
Sequence: 1 MKNFNQYFITLISMLVAC.....GAQFSETNCRHVLKPKNL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	6.5	320	1 YCFS_ECOLI	P75954 escherichia
2	84	6.2	599	1 VG58_HSV11	Q00157 ictaluriid h
3	83	6.1	209	1 L0LB_HAEIN	P45270 haemophilus
4	83	6.1	528	1 YAC8_YEAST	P39734 saccharomyc
5	82.5	6.1	297	1 SYGA_BACHD	Q9kd49 bacillus ha
6	82.5	6.1	365	1 MLTA_ECOLI	P46885 escherichia
7	80.5	5.9	474	1 PEDLA_LACHE	Q48558 lactobacill
8	80	5.9	369	1 CD5S_MOUSE	Q35926 mus musculu
9	77.5	5.7	594	1 NIR_SPIOL	P05314 spinacia ol
10	77	5.7	350	1 REDD_STRCO	P16922 streptomyce
11	77	5.7	506	1 VLL_BPVA	P08341 bovine papi
12	77	5.7	1107	1 ALA2_ARATH	P98205 arabidopsis
13	76.5	5.6	316	1 COAA_BACHD	Q9kx77 bacillus ba
14	76.5	5.6	852	1 RBMA_RAT	P70501 rattus norv
15	76	5.6	865	1 MUTS_BACHD	Q9kac0 bacillus ha
16	75.5	5.6	390	1 VASS_BPGA	P07394 bacterioph
17	75.5	5.6	563	1 PHNL_DESFR	P18188 desulfovibr
18	75	5.5	321	1 CATO_HUMAN	P43234 homo sapien
19	75	5.5	368	1 ALR_LISMO	O85045 listeria mo
20	75	5.5	746	1 YUYI_CAEEL	P54073 caenorhabdi
21	74.5	5.5	467	1 VNSS_TSWVL	P26003 tomato spot
22	74.5	5.5	912	1 KPCM_HUMAN	Q15139 homo sapien
23	74	5.5	438	1 XYLE_THEYO	Q9kgu2 thermoanaer
24	74	5.5	1053	1 CAPP_SYN6	P06516 synecococc
25	74	5.5	1189	1 PTNE_MOUSE	Q62130 mus musculu
26	74	5.5	2190	1 CCAD_CHICK	O73700 gallus gall
27	74	5.5	2210	1 RRPL_EBOSM	Q66802 ebola virus
28	73.5	5.4	449	1 TRB2_AERPE	Q9y9h2 aetopyrum p
29	73.5	5.4	595	1 YE72_HUMAN	Q9nx95 homo sapien
30	73.5	5.4	615	1 CPO_DRONE	Q01617 drosophila
31	73.5	5.4	1440	1 Y8EO_PSEAE	P14403 j genome po
32	73	5.4	212	1 Y8EO_PSEAE	Q9hvt1 pseudomonas
33	73	5.4	448	1 XYLE_BACLI	P77832 bacillus li

ALIGNMENTS

```

RESULT 1
YCFS_ECOLI
AC P75954; STANDARD; PRT; 320 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycf5 precursor.
GN YCFS OR B1113
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE ERPK/YBIS/YCFS/YNHG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000211; AAC74197.1; -
CC EMBL: D90746; BAA35928.1; -
CC EMBL: D90747; BAA35933.1; -
CC EcoGene: EG13437; ycf5.
CC InterPro: IPR002482; LysM.
CC Pfam: PF01476; LysM; 1
CC SMART: SM00257; LysM; 1
CC Hypothetical protein; Periplasmic; Signal; Complete proteome.
CC SIGNAL 1 23
CC POTENTIAL.

```

P49803 rattus norv
P49802 homo sapien
P10394 drosophila
Q10135 schizosacch
P20806 drosophila
Q9nrc6 homo sapien
Q81023 human papil
P26191 porcine rot
O84417 chlamydia t
P28955 equine herp
P53068 saccharomyc
Q59940 streptococc

FT CHAIN 24 320 HYPOTHETICAL PROTEIN YCFS.
 SQ SEQUENCE 320 AA; 34636 MW; 89378ABDFD60359B CRC64;

Query Match
 Best Local Similarity 6.5%; Score 88; DB 1; Length 320;
 Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITLSSMLVACSAPIPNQVSPKTPSLTKDKIGDHTHEDESVSHVGLQAHE 67
 DB 10 WLTFEFAAVALPAKATWPLP-PAGSRVGENKF--HVVDGG----- 54
 QY 68 TWLQMHATKQEVRYQAYLQSLRG--NYLPPMSQLLTARSQACGHEPYOLPEHLWG 125
 DB 55 ---SLEAIKKYVNGFLALLQANPGVDVYPRAGSVLT----- 89
 QY 126 QYPTTHLYODLSKRGILPANTOIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176
 DB 90 --IPLQTLPLDAPREGIVINAEIRLYYPPGKNSVTYPIGIGQLGGDTLTPTWTVTVS 147
 QY 177 D-----IWPVDEIKSQAELYQN-----RLCQY---WLEHGENQ 208
 DB 148 DKRANPTWTANIR-ARYKAQGIELPVPVAGLDPNMGHHAIRLAAVGGVYLLHGTNA 205
 QY 209 NFGGL-L-YATGAHL 222
 DB 206 DFGIGMRVSSGICRL 220

RESULT 2
 VG58_HSVII STANDARD; PRT; 599 AA.
 AC Q00157;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE Hypothetical gene 58 protein.
 GN 58.
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Unclassified Herpesviridae.
 OX NCBI_TaxID=10401;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AUBURN 1;
 RX MEDLINE=92087490; PubMed=1727613;
 RA Davison A.J.;
 RT "Channel catfish virus: a new type of herpesvirus."
 RL Virology 186:9-14(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M75136; AAA88161.1; .
 DR PIR; E36792; E36792.
 KW Hypothetical protein.
 SQ SEQUENCE 599 AA; 67476 MW; 991C2B524F0F6BEF CRC64;

Query Match
 Best Local Similarity 6.2%; Score 84; DB 1; Length 599;
 Matches 58; Conservative 50; Mismatches 95; Indels 116; Gaps 15;

QY 1 MKNFNQVFTTLISML-----VACSAPIPTNPQ--VSPFKTPSVLITKDKIGDHTH 51
 DB 267 MAMFRYPIDRLERIMDMYFGSGLKCTLATFTTPKTYVSGSRHSIVIKNHVDKHYLN 326
 QY 52 EHDESVSHVGLQAHEFTWLQMHATKQEVRYQAYLQSLRGVLYNLPMSQLLTARSQAC 111

DB 327 DNSKFLAHVMDRAMPTTCYIH-----DDIDVKAMIQSVISRVVVRMIQDTQVRLQELSA 381
 QY 112 GHEPY-----OLPPE-----HLWGQI-VPTL-----HLYQDL--KSRG 141
 DB 392 GNKLHFHFFNQLPPPMNNLDISFKYNPLVDHALQGVPGVGLQYGNPEQIYTDMITEMTA 441
 QY 142 ILPA-----NTQIRS-----VYRN-----PELNQACGAAM----- 167
 DB 442 LLPRIGHMVPEMWSQELSSSKTKLCLLPKIARDLNPVPLVTSLNDDTGNAMLJAFHAKI 501
 QY 168 -----SKHLNSAIDIWVPDEIKSQAELYQN----- 195
 DB 502 YIATAMLIHLHCENLETHDLGPOTGTGVGPEPKHKTGAQLREWFIDLQ-KTIAL-DAPS 559
 QY 196 RLC-----QYWLEHGENQNF 210
 DB 560 THCACCADFWLTHGSDPNF 578

RESULT 3
 LOLB_HAEIN STANDARD; PRT; 209 AA.
 AC P45270;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer-membrane lipoprotein lolo precursor.
 GN LOLB OR HEMM OR H1607.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fierschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerslaghe A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd."
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: PLAYS A CRITICAL ROLE IN THE INCORPORATION OF
 CC LIPOPROTEINS IN THE OUTER MEMBRANE AFTER THEY ARE RELEASED BY
 CC THE LOLA PROTEIN (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE LOLB FAMILY.
 CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE INVOLVED IN DELTA-
 CC AMINOLEVULINIC ACID BIOSYNTHESIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32834; AAC23251.1; .
 DR TIGR; H1607; .
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Chaperone; Outer membrane; Lipoprotein; Transport; Protein transport;
 KW Signal; Complete proteome.
 FT SIGNAL 1 21 BY SIMILARITY.
 FT CHAIN 22 209 OUTER-MEMBRANE LIPOPROTEIN LOLB.

```
FT LIPID 22 22 N-ACYL DIGLYCERIDE (BY SIMILARITY).
SQ SEQUENCE 209 AA; 24193 MW; 0555F02F13E852A1 CRC64;

Query Match 6.1%; Score 83; DB 1; Length 209;
Best Local Similarity 23.0%; Pred. No. 1.7;
Matches 59; Conservative 36; Mismatches 87; Indels 74; Gaps 17;

QY 1 MKNFQYFITTLLISSMLVACSAPI--PTNPQVSPKTPSVLITKDKIGDHTTHEHDSVS 58
Db 4 MKTF-KFTALFAITALLDMEPTNPVQYID-KTDAI----- 41
QY 59 HVGLOAHFETWLOMHATKQEVRYQAYLQSRGLNGLPMSOLLTTARSQACGHEPYQL 118
Db 42 -----W-QQHLQIKIQSYQA--KGQIG-YISP-TERFSRFEWQ-----YQN 80
QY 119 PPEHLWGOIVTLLHYQDLKSRGILPANTQIRSVYRNPELNQCAGGAAKSHLTNSAIDI 178
Db 81 PKSY-----TLKLSYLSIKSTLWQHQSGMTISDNNGNQ--QSAANSKLILQEIIGM 131
QY 179 WVPDLTKSQALYELQNLRCQYWL--EHEGENONFGLGL-VATGAIHLDTGFRKWAQ-- 233
Db 132 DVP-----LEH--LAYWLKGPAMNADYQVGTNHLGAFTHVDG--SOWTADYL 177
QY 234 -FSETNSICRHVLPKN 248
Db 178 TYHSNMPENILKN 193

RESULT 4
YAC8_YEAST STANDARD; PRT; 528 AA.
AC P39734; P39733;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 58.8 kDa protein in MYO4-DRS2 intergenic region.
GN YAL028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [2]
RP REVISIONS.
RA Vo D.T.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U12980; AAC05004.1; -
CC SGD; S0000026; YAL028W.
CC KW Hypothetical protein; Transmembrane.
CC DOMAIN 308 328 POLY-SER.
CC FT TRANSMEM 505 521 POTENTIAL.
CC FT DOMAIN 512 517 POLY-ILE.
CC SEQUENCE 528 AA; 58750 MW; E5A3CC7C6D60977A CRC64;

Query Match 6.1%; Score 83; DB 1; Length 528;
Best Local Similarity 19.4%; Pred. No. 5.2;
Matches 41; Conservative 35; Mismatches 89; Indels 45; Gaps 6;

QY 3 NFQYFITTLLISSMLVACSAPIPTNPQVSPKTPSVLITKDKIGDHTTHEHDSVSHVGL 62
Db 308 SFSQSSSSSSSSSSSSSTFSQSVAVDPLEPPGNITYSSSNLSNLSNDELQYQRIHGL 367
QY 63 QAHFETWLOMHATKQEVRYQAYLQSRGLNY--LPPMSOLLTTARSQACGHEPYQL- 118
Db 368 QLQ-QTEALLKHSKDEVLDKENDLVKNTANFDKIVKELDRSRFTIGWKELVEDYLMN 426
QY 119 -----PPEHLWGOIVTLLHL-----YQDLKSRGILPANTQIRSVYRNPELNQCAGGAA 166
Db 427 LKODFDKENPESPEARLSDTINTNVAKLQDLEKR-----MASCKDRLA 469
QY 167 MSKHLTNSAIDIWVPDLTKSQALYELQNLRL 197
Db 470 SRKEVMR-----KMESLLSLENSL 488

RESULT 5
SYGA_BACHD STANDARD; PRT; 297 AA.
AC O9KD49;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycyl-tRNA synthetase alpha chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE alpha chain) (GLYRS).
GN GLYO OR BH1370.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AP001511; BAB05089.1; -
CC InterPro; IPR002106; AA_trna_ligase_II.
CC InterPro; IPR002310; trna_synt_2e.
CC Pfam; PF02091; trna_synt_2e; 1.
CC PRINTS; PR01044; TRNASYNTHGA.
CC PROSITE; PS00179; AA-trna_ligase_II_1; FALSE_NEG.
CC PROSITE; PS00339; AA-trna_ligase_II_2; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SEQUENCE 297 AA; 34420 MW; F27D6C2C74969E7 CRC64;
```

```

Query Match          6.1%; Score 82.5; DB 1; Length 297;
Best Local Similarity 24.3%; Pred. No. 2.8; 56; Indels 59; Gaps 12;
Matches 44; Conservative 22; Mismatches 55;

QY 67 ETWLQMHATKQEVRYQAYLQSLRGLNPLPPMSQLLTTA-RSQWACGHEPYQLPPEHLWG 125
   ||| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB 12 EYW-----SKQNCILLQAYDEKAGTMSPTMLRTTGPENWVAYVEPSRRPADRGY 65

QY 126 -----QIV-----PT--LHLYQD-LKSRGILPANTOIRSV---YRNPENLQACAG 163
   ||| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 ENPNRLYQHOFQHKMPSFTNIQIYLDLSRLGALGNPLEHIVDENWENPSLG-CAG 124

QY 164 GAAMSKHLNTSAIDIWVPLEI-----KSOALYELQNLRCQYWLHGE 206
   ||| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 -----LQWEWLDGMEITQFTYFOQVGLGANPVSABITGLE-RLASY-IDQKE 172

QY 207 N 207
   |
DB 173 N 173

RESULT 6
MLTA_ECOLI STANDARD; PRT; 365 AA.
AC P46885; P76638; Q46928;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-)
DE (Murein hydrolase A) (Mlt38).
GN MLTA OR MLT OR B2813 OR Z4130 OR ECS3673.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX STRAIN=K12;
RX MEDLINE=97431497; PubMed=9287002;
RA Lommatzsch J., Templin M.F., Kraft A.R., Vollmer W., Hoeltje J.-V.;
RT "Outer membrane localization of murein hydrolases: MltA, a third
RL lipoprotein lytic transglycosylase in Escherichia coli.";
RL J. Bacteriol. 179:5465-5470(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

```

```

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [5]
RP SEQUENCE OF 147-161; 204-213 AND 258-280, AND CHARACTERIZATION.
RX MEDLINE=94117367; PubMed=8288527;
RA Ursinus A., Hoeltje J.-V.;
RT "Purification and properties of a membrane-bound lytic
RT transglycosylase from Escherichia coli.";
RL J. Bacteriol. 176:338-343(1994).
RN [6]
RP INTERACTION WITH MIPA AND MRCP/PONB.
RX STRAIN=ATCC 53338 / MC1061;
RX MEDLINE=99156961; PubMed=10037771;
RA Vollmer W., von Rechenberg M., Hoeltje J.-V.;
RT "Demonstration of molecular interactions between the murein polymerase
RT PBP1B, the lytic transglycosylase MltA, and the scaffolding protein
RT Mipa of Escherichia coli.";
RL J. Biol. Chem. 274:6726-6734(1999).
CC -!- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING OF
CC MUROPEPTIDES DURING CELL ELONGATION AND/OR CELL DIVISION. OPTIMAL
CC ACTIVITY IS BETWEEN PH 4.0 AND 4.5; LOSES ITS ACTIVITY RAPIDLY AT
CC TEMPERATURES ABOVE 30 DEGREES CELSIUS. DEGRADES MUREIN GLYCAN
CC STRANDS AND INSOLUBLE, HIGH-MOLECULAR WEIGHT MUREIN SACCULI.
CC -!- CATALYTIC ACTIVITY: CLEAVAGE OF THE BETA-1,4-GLYCOSIDIC BOND
CC BETWEEN N-ACETYLGLUCOSAMINE AND N-ACETYLGLUCOSAMINE RESIDUES,
CC THEREBY CONSERVING THE ENERGY IN A NEWLY SYNTHESIZED
CC 1,6-ANHYDROBOND IN THE MURAMIC ACID RESIDUE.
CC -!- SUBUNIT: FORMS A TRIMERIC COMPLEX WITH MRCP/PONB AND MIPA IN
CC VITRO.
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U32224; AAC45723.1; -
CC EMBL; AE000365; AAC75855.1; -
CC EMBL; U29581; AAB40463.1; ALT_INIT.
CC EMBL; AE005509; AAG57927.1; -
CC EMBL; AP002563; BAB37096.1; -
CC EcoGene; EG13085; mltA.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC Cell wall; Hydrolase; Glycosidase; Signal; Lipoprotein;
KW Outer membrane; Multigene family; Complete proteome.
FT SIGNAL 1 20 PROBABLE
FT CHAIN 21 365 MEMBRANE-BOUND LYTIC MUREIN
FT LIPID 21 21 TRANSGLYCOSYLASE A.
FT CONFLICT 346 346 H -> N (IN REF. 1).
FT SEQUENCE 365 AA; 40410 MW; 5ECB92C1E8D5969 CRC64;
SQ

```

```

Query Match          6.1%; Score 82.5; DB 1; Length 365;
Best Local Similarity 19.1%; Pred. No. 3.7;
Matches 57; Conservative 43; Mismatches 88; Indels 111; Gaps 16;

QY 6 QYFITLSSMLVACSA-PIPTNPQVSPK--TPSVLITK-DKIG-DHRTHEDESVSHV 60
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 7 KYLLMGTVVAMLAACSSKPTDRGQYKDGKFTQPSFLVNQPDVAGCAPINAGDAEQINHI 66
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 61 -----GLOAH-----FETWL-----QMHATKQEVRYQAYLQSLR 91
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 67 RNSSPRLYGNQSNVNAVQEWLRAGGDTNRMRQFCIDAWQMEGADNYGNVOTGY----- 121
QY 92 GNYLPPMSOLLTARSQWACGHEPYQLPPEHLWGQIVFTLHLYQDLKSRGILPANTQIRS 151
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```


Db 122 --YTFVIOARHTROGEFQ---YPIYRMPK-----RGRLPRAEIIYA 158
QY 152 VYRNELNQCAGGAMSKHL---TNSAIDIVWPDLEIKSQALYELQNLRCQYWEHGENQ 208
Db 159 -----GALSDRYIIAYNSLMDNFIMDVQSG-----YIDFGDGS 193
QY 209 NFGGLGYA-----TGAIHLDD-----TQGFRRKMGACFSETNSICRHVLPKN 248
Db 194 PLNFFSYAGKNGHAYRSIGKVLIDRGEVKKEDMSQAIIRHWGTHSEAE--VRELLLEQN 250

RESULT 7
PEDA_LACHE
ID PEDALACHE STANDARD; PRT; 474 AA.
AC Q48558; P71434;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dipeptidase A (EC 3.4.-.-).
GN PEDPA OR PEDP.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CMR232;
RX MEDLINE=96146518; PubMed=8550503;
RA Dudley E.G., Husegen A.C., He W., Steele J.L.;
RT "Sequencing, distribution, and inactivation of the dipeptidase A gene
(pepDA) from Lactobacillus helveticus CNR232.";
RL J. Bacteriol. 178:701-704(1996).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=53/7;
RX MEDLINE=96338998; PubMed=8766699;
RA Vesanto E., Peltonlehti K., Purtsi T., Steele J.L., Palva A.;
RT "Molecular characterization, over-expression and purification of a
novel dipeptidase from Lactobacillus helveticus";
RL Appl. Microbiol. Biotechnol. 45:638-645(1996).
CC -!- FUNCTION: HYDROLYZES A WIDE RANGE OF DIPEPTIDES BUT UNABLE TO
CC HYDROLYZE DIPEPTIDES CONTAINING PROLINE. HIGHEST ACTIVITY AGAINST
CC MET-ALA.
CC -!- CATALYTIC ACTIVITY: DIPEPTIDE + H(2)O = 2 AMINO ACID.
CC -!- ENZYME REGULATION: INHIBITED BY ZN(2+), CU(2+), CA(2+) AND CD(2+).
CC -!- SUBUNIT: HOMOCOTAMER (PROBABLE).
CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS OBSERVED AT PH 6.0 AND 55
CC DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY U34.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U34257; AAC43971.1; -.
DR EMBL; Z38063; CAA86210.1; -.
DR MEROPS; U34.001; -.
KW Hydrolase; Dipeptidase.
FT CONFLICT 211 211 T -> A (IN REF. 2).
SQ SEQUENCE 474 AA; 53512 MW; 3BFA79983D3CEEF2 CRC64;

Query Match 5.9%; Score 80.5; DB 1; Length 474;
Best Local Similarity 23.2%; Pred. No. 7.6;
Matches 39; Conservative 28; Mismatches 56; Indels 45; Gaps 9;
QY 43 DKIGDHHTHDESVSHVGLQAFETW-LQMHHATKQEV--VRYOAYLQSLGNVLPNMS 99
Db 301 DAYGDQGTPEQKTFPIGINRRNFETHILQIRNDVPAETAGVQWLAFGPNFTNSMLPFYT 360

QY 100 QLTTARTSWOACGHEPYQLPPEHL---W-----GOIVPT-LHLYQDLKSRGILPANTOI 149
Db 361 NVTTTPEAWQT-----TPKFNLIKFWLNKLTAGLDGTDNRYVGELE-----402
QY 150 RSVYRNPELNQCAGGAMSKHLTNSAIDIVWPDLEIKSQALYELQNLRL 197
Db 403 -DAFEOKSLAQC-----HKIQHET-----DKEYKNLSGRELQDKL 436

RESULT 8
CD5S_MOUSE
ID CD5S_MOUSE STANDARD; PRT; 369 AA.
AC Q35926; O35277;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cyclin-dependent kinase 5 activator 2 precursor (CDK5 activator 2)
DE (Cyclin-dependent kinase 5 regulatory subunit 2) (P39) (P391).
GN CDK5R2 OR NCKSAI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98322126; PubMed=9655938;
RA Nilden F., Baekstrom A., Bark C.;
RT "Molecular cloning and characterisation of a mouse gene encoding an
isoform of the neuronal cyclin-dependent kinase 5 (CDK5) activator.";
RL Biochim. Biophys. Acta 1398:371-376(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Zheng M., Leung C.L., Liem R.K.H.;
RT "Comparative analysis of gene expression of the cyclin-dependent
Kinase 5 (cdk5) activators p35 and p39 in the rat CNS and embryonic
development revealed distinctive overlap with cdk5.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTIVATOR OF CDK5/TPKII.
CC -!- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U90267; AAC53595.1; -.
DR EMBL; AF016393; AAB69709.1; -.
DR MGD; MGI:1330828; Cdk5r2.
FT PROPEP 1 ?
FT CHAIN ? 369 ?
FT DOMAIN 75 78 POTENTIAL.
FT DOMAIN 155 163 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 348 351 POLY-LYS.
FT DOMAIN 348 351 POLY-PRO.
FT CONFLICT 91 91 G -> V (IN REF. 2).
FT CONFLICT 172 172 S -> G (IN REF. 2).
FT CONFLICT 191 191 R -> G (IN REF. 2).
FT CONFLICT 197 197 S -> R (IN REF. 2).
FT CONFLICT 220 220 L -> W (IN REF. 2).
FT CONFLICT 248 248 S -> R (IN REF. 2).
FT CONFLICT 253 253 A -> G (IN REF. 2).
SQ SEQUENCE 369 AA; 38914 MW; DA4EB1FD4D48CE66 CRC64;

Query Match 5.9%; Score 80; DB 1; Length 369;
Best Local Similarity 27.5%; Pred. No. 6.2;

Heme; Iron-sulfur; 4Fe-4S.

```

KW TRANSIT 1 32 CHLOROPLAST.
FT CHAIN 33 594 FERREDOXIN--NITRITE REDUCTASE.
FT METAL 473 473 IRON-SULFUR OR HEME IRON (POTENTIAL).
FT METAL 479 479 IRON-SULFUR OR HEME IRON (POTENTIAL).
FT METAL 514 514 IRON-SULFUR OR HEME IRON (POTENTIAL).
FT METAL 518 518 IRON-SULFUR (4FE-4S) AND SIROHEME
FT VARIAT 221 221 (BY SIMILARITY).
FT SEQUENCE 594 AA; 66394 MW; 2B3DCAAC16DE06A3 CRC64;

Query Match 5.7%; Score 77.5; DB 1; Length 594;
Best Local Similarity 22.0%; Pred. No. 19;
Matches 63; Conservative 35; Mismatches 97; Indels 91; Gaps 13;

QY 27 NP-QVSPKTPSVLTKKIGDHHT-----HEHDESVSHVGLQAHFTWLMHAT 76
   || || || || || || || || || || || || || || || || || || ||
Db 77 NPAEKVIEKDPMKLFIEDGISDLATLSMEVDKSHNKDDIDVRLK-----WLGFLFHR 131
   || || || || || || || || || || || || || || || || || || ||
QY 77 KOEVRVQY-----AYLOSRLGNVLPPMSOLLTTLARSQWACGHEPPYQLPPE 121
   || || || || || || || || || || || || || || || || || || ||
Db 132 KHWYGRFMRMLKLPNGVTTSEQTRYLASVTKKGKCCADVTTQNWQ----- 179
   || || || || || || || || || || || || || || || || || || ||
QY 122 HLMGQIVPTL-HLYQDLKSRGILPANTQIRSVYRNPELNQCAG-----GAAMS 168
   || || || || || || || || || || || || || || || || || || ||
Db 180 -IRGVLPDPVEIKLGESVGLTSLQSGMDNV-RNPVGNPLAGIDPHEIVDTVPFTNLIS 237
   || || || || || || || || || || || || || || || || || || ||
QY 169 KHLT-NSAIDIWPDLE-----IKSQALYEIQNLRCQYWLEHGENQFGLGYATG-- 218
   || || || || || || || || || || || || || || || || || || ||
Db 238 QFVTRANSRGLSTNLPKWNPCVIGSHDIYHPHINDLAYMPATKNGKGFNLLVGGFF 297
   || || || || || || || || || || || || || || || || || || ||
QY 219 -----AIHLDT-----QGFRKWGAQFSETNSICR 242
   || || || || || || || || || || || || || || || || || || ||
Db 298 SIKRCEAIPLDAWSAEDVVPVCKAMLEAFRLDG--FRGNRQKCR 341
   || || || || || || || || || || || || || || || || || || ||

RESULT 10
REDD_STRCO ID REDD_STRCO STANDARD; PRT; 350 AA.
AC PL6922;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Transcriptional regulator redd.
GN REDD OR SC2E9.18.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid:1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=90094237; PubMed=2294088;
RA Narva K.E., Feltelson J.S.;
RT "Nucleotide sequence and transcriptional analysis of the redd locus
   of Streptomyces coelicolor A3(2).";
RL J. Bacteriol. 172:326-333(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
PL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE REDD PROTEIN IS PROBABLY ONE OF SEVERAL DELICATELY
CC BALANCED REGULATORY FACTORS THAT CONTRIBUTE TO THE CONTROL OF
CC THE BIOSYNTHESIS OF THE ANTIBIOTIC UNDECYLPRODIGISIN (RED) IN
CC S.COELICOLOR.
CC -!- SIMILARITY: BELONGS TO THE AFSPR/DNRI/REDD FAMILY OF REGULATORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```



```

FT TRANSMEM 808 827 POTENTIAL.
FT DOMAIN 828 841 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 842 860 POTENTIAL.
FT DOMAIN 861 890 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 891 912 POTENTIAL.
FT DOMAIN 913 919 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 920 942 POTENTIAL.
FT DOMAIN 943 948 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 949 969 POTENTIAL.
FT DOMAIN 970 982 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 983 1007 POTENTIAL.
FT DOMAIN 1008 1107 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 381 391 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 752 752 MAGNESIUM (BY SIMILARITY).
FT METAL 756 756 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1107 AA; 124835 MW; 31F7729E9653C96F CRC64;

Query Match 5.7%; Score 77; DB 1; Length 1107;
Best Local Similarity 19.2%; Pred. No. 45;
Matches 59; Conservative 40; Mismatches 85; Indels 124; Gaps 15;

QY 5 NOYFITLISMLVAC-----SAPIPNPQVSPKTPSVLI-----TKKIGDHHTHEHDE 55
Db 46 NOYF-----LLIACLQLMSLITPNP--ASTWGPLIFAFVASKAEAWDDYHRYLSDK 96
QY 56 SVSHVGLQAHFETWLMHHTAKOEVVRYQAYLOSRLGNY-----LPPMSQLLTARS 107
Db 97 KANEK-----EWI-----VRQGIKKHQAQDIQVGNIVLWLENDEVCPLVLTGSDP 145
QY 108 WQACGHEPYQLPPEHLWGIVPTLHLYQDLKSRGILPANTQIRSVYRNPENLQACGAGAM 167
Db 146 QGVYVETAALDGE-----TDLKTR-VIP-----SACVGIDLE 177
QY 168 SKHLTNSAIDWIPDLEIK-----SQALYE--LQNRLCQYWLHEGENON-----FGL 212
Db 178 LLHKMGVIECPVPDKDIRREDANMLRFPFDINDVCSLTIRKNTLLQSCYLYRNTEWACGV 237
QY 213 GLVA-----TGAIHL-----DQGFPR 228
Db 238 SVYTGNTKLGMSRGIAEPKLTAMAMIDKLTGAEIVQVIVVVLVGLIAGNVKWKDEARK 297
QY 229 KWGAQFSE 236
Db 298 QWYVQYPE 305

RESULT 13
COAA_BACHD STANDARD; PRT; 316 AA.
AC Q9K8X7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).
GN COAA OR BH2875.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
CC phosphopantothenate.
CC -!- PATHWAY: Coenzyme A (COA) biosynthesis; first step.

```

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE PANTOTHENATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001516; BAB06594.1; -.
CC InterPro; IPR001324; PRK.
CC Pfam; PF00485; PRK; 1.
CC Transferrase; Kinase; ATP-binding; Coenzyme A biosynthesis;
CC Complete proteome.
CC NP_BIND 95 102 ATP (POTENTIAL).
CC FT NP_BIND 316 AA; 36416 MW; 3196BA013E3B0BB3 CRC64;
CC SQ SEQUENCE 316 AA; 36416 MW; 3196BA013E3B0BB3 CRC64;

Query Match 5.6%; Score 76.5; DB 1; Length 316;
Best Local Similarity 26.4%; Pred. No. 11;
Matches 32; Conservative 20; Mismatches 42; Indels 27; Gaps 7;

QY 109 QACGHEPYQLPPEHLWGIVPTLHLYQDLKSRGILPANTQIRSVYRNPENLQACGAGAMS 168
Db 5 EACDFFPYTVLSRSQW-----KSLRKASSLPINEQ-----ELEQLVG---LN 43
QY 169 KHLT-NSAIDTWVP---DLEIKSQALYELQNRLCQYWLHEGENON-FGLGIYATGAIHLD 223
Db 44 EPTILNEVADIYVPLAELLHVHATAYQLQOQK-RGFHHGKNRSPFIIGLAGSVAVGKS 102
QY 224 T 224
Db 103 T 103

RESULT 14
REMA_RAT STANDARD; PRT; 852 AA.
AC P70501;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE RNA-binding protein 10 (RNA binding motif protein 10) (sl-1 protein).
GN RBM10.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96357133; PubMed=8760884;
RA Inoue A., Takahashi K., Kimura M., Watanabe T., Morisawa S.;
RT "Molecular cloning of a RNA binding protein, sl-1.";
RL Nucleic Acids Res. 24:2990-2997(1996).
CC -!- FUNCTION: NOT KNOWN. BINDS TO RNA HOMOPOLYMERS, WITH A PREFERENCE
CC FOR POLY(G) AND POLY(U) AND LITTLE FOR POLY(A).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D83948; BAA12144.1; -.

```



```

Db 115 NRELWNAITPLRLVLYLVAAIL-TDFEVTSVYRDLPLNQAGGANSRHLFNSAIDFR 173
QY 180 V-PDLEIKSQALYELQN---RLCYWLEHGENONFGLGYATGAIHLDTQGRKWAQFS 235
Db 174 IGEIP-QQDFAFIENTFKLCQFNQHGQSLNGLGLYSQGIHDTQGYRTWGPDS 232
QY 236 ETSNCRH 243
Db 233 RNSSMCNY 240
RESULT 2
Q92JX1 PRELIMINARY; PRT; 562 AA.
AC Q92JX1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SMC04010.
GN SMC04010.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=1021;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
Barloy-Hubier F., Barnett M.J., Becker A., Boistard P., Botie G.,
Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
Glox S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
Ramsperger U., Surzycki R., Thebault P., Vandenbol M., Vandenbol M.,
Vorhoeelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti."
RL Science 293:668-672(2001).
EMBL: AL591791; CAC47373.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 562 AA; 59535 MW; 12C41231BACD3422 CRC64;

Query Match 9.0%; Score 122; DB 16; Length 562;
Best Local Similarity 25.4%; Pred. No. 0.0024;
Matches 43; Conservative 29; Mismatches 55; Indels 42; Gaps 10;

QY 74 HATKQEVVRYQAYLQSLRGNY-LPPMSQLLTARSQWACGHEPYQLPPEHL---WQIYP 129
Db 20 HTKEKQITKY-----RNGRYDQKGLQINRFLRDWRR--NEPTKMDPRLDLVW----- 67
QY 130 TLHLQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLE 184
Db 68 --EYQKSGRDI-----HVSAYRSPATNGMLRSKRGVAKSKQHMLGKAMDFIYDVK 121
QY 185 IKSQALYELQNLCQWLEHGENONFGLGYATGA---IHLDTQGRKW 230
Db 122 LKT--LREVGKMF-----QVGGVGYPTSGSPFVHMDVGVRAW 158

RESULT 3
Q98H74 PRELIMINARY; PRT; 622 AA.
AC Q98H74;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE ML2999 PROTEIN.
GN ML2999.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```

```

OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
EMBL: AP003000; BAB49992.1; -.
KW Complete proteome.
SQ SEQUENCE 622 AA; 65402 MW; 1E4C6B5F8C291655 CRC64;

Query Match 7.9%; Score 106.5; DB 16; Length 622;
Best Local Similarity 23.4%; Pred. No. 0.086;
Matches 40; Conservative 27; Mismatches 61; Indels 43; Gaps 9;

QY 72 MHATKQEVVRYQAYLQSLRGNYLPP-MSQLLTARSQWACGHEPYQLPPEHL---WGQI 127
Db 27 LHTEKAEIV-----YKRNRYLPEGLRKINILRDWRR--NEPTKMDPRLDLVW--- 75
QY 128 VPTLHLQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDP 182
Db 76 ---EAYRESGATDYI---QVVCYRSPATNSMLRSKRGVAKSKQHMLGKAMDFIYIPG 127
QY 183 LEIKSQALYELQNLCQWLEHGENONFGLGYATGA---IHLDTQGRKW 230
Db 128 VPLKLRNIGL-----KMGGVGYPTSGSPFVHMDVGVNRHW 166

RESULT 4
Q989B8 PRELIMINARY; PRT; 523 AA.
ID Q989B8
AC Q989B8;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE MLR6494 PROTEIN.
GN MLR6494.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
EMBL: AP003009; BAB52779.1; -.
KW Complete proteome.
SQ SEQUENCE 523 AA; 55318 MW; FA6196E5665414C4 CRC64;

Query Match 7.6%; Score 102.5; DB 16; Length 523;
Best Local Similarity 25.4%; Pred. No. 0.17;
Matches 43; Conservative 29; Mismatches 58; Indels 39; Gaps 10;

QY 72 MHATKQEVV--RYQAYLQSLRGNYLPPMSQLLTARSQWACGHEPYQLPPEHLWGQIYP 129
Db 28 LHTGEKAEIVKRNRYDQAGL-----KKIDFMLRDWRR--NEPTKMDPRLI--DLV- 75

```


Query Match	7.5%	Score 101;	DB 16;	Length 216;
Best Local Similarity	22.7%	Pred. No. 0.073;		
Matches	50;	Conservative	79;	Indels 62; Gaps 10;
QY	28	POVSPKTPSVLIT-----KDKIGDHHTHDEDSVHVGLOAHFETWLMQHHATK	77	
Db	39	POATVPVPTTVAATVASIDPPALKPAVDPRWVHLN-----VHTGEK	81	
QY	78	QEVRYQAYLOSRLGNKYL-PMSELLTATRSWACGHEPQLPPEHLWGQIVPTLHLQYD	136	
Db	82	LEAVYWN-----GDVYDVAVSALDKVLRDYRDEVHPIDRCLYDLDDQIA-----RK	129	
QY	137	LKSRGILPANTQIRSVYRNPELNCAG-----AAMSKHLTNSAIDWIWPDLEIKSQALY	191	
Db	130	TQSKGPF-----QVTSIGYKSPAINLLSKRSGEYAKKSLHMDGKAMDIFLEDVELXHVRAA	185	
QY	192	ELQNLRCQYWLHEHGENOFGLYAT-GAHLDTQGRKW	230	
Db	186	ALD-----LSVG-----GVGYPTSFVHVDVGPVRKW	213	
RESULT	6			
Q98A17		PRELIMINARY;	PRT;	193 AA.
ID	Q98A17			
AC	Q98A17;			
DT	01-OCT-2001 (TrEMBLrel. 18, Created)			
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)			
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)			
DE	ML15985 PROTEIN.			
GN	ML15985.			
OS	Rhizobium loti (Mesorhizobium loti).			

```

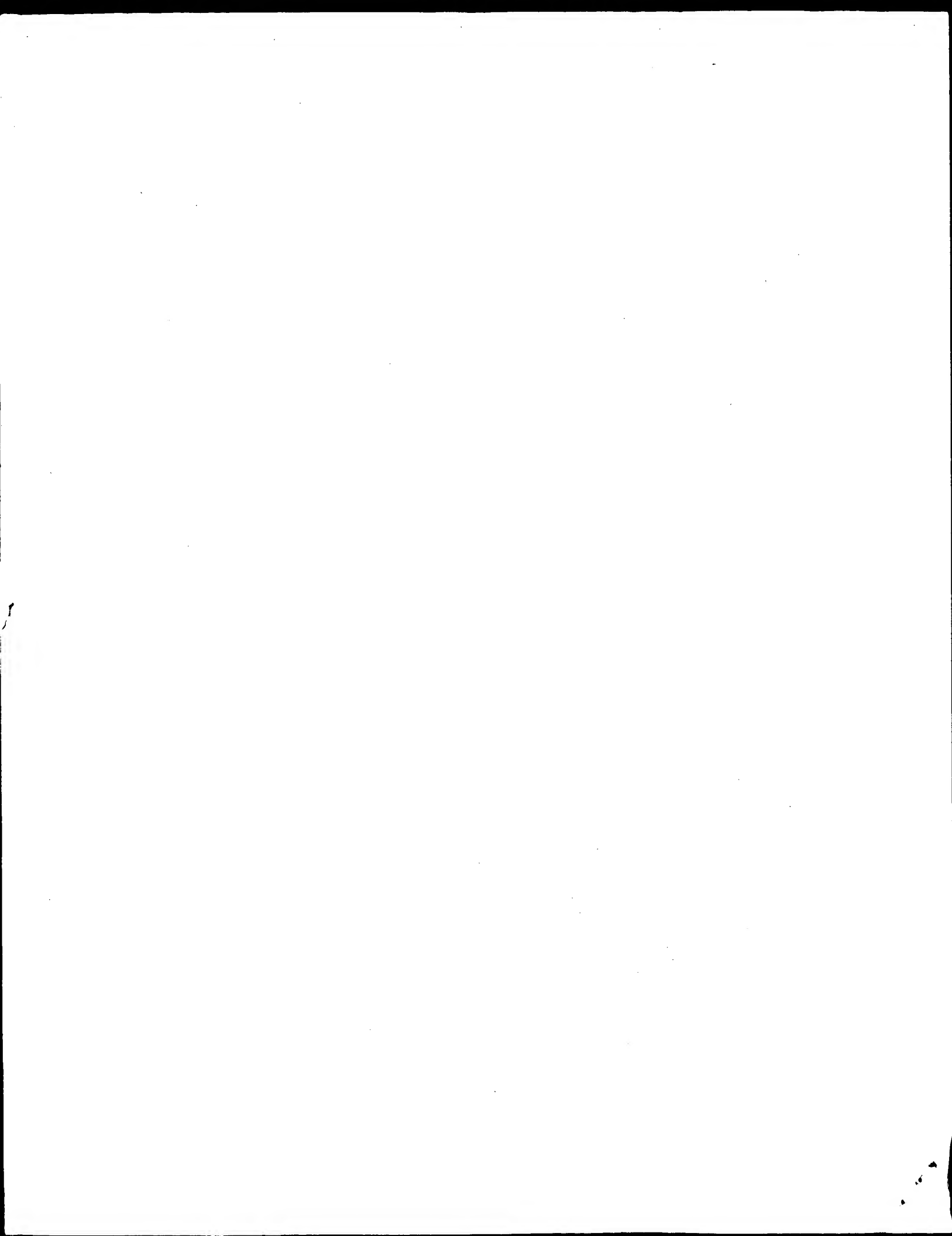
Query Match          6.68; Score 90; DB 5; Length 539;
Best Local Similarity 27.2%; Pred.No. 2.8;
Matches 41; Conservative 19; Mismatches 63; Indels 28; Gaps 6;

QY   54  DESVSHVGLQAHFETWQMHHATKQEVVRQVAQLQSRLGNYLPPMSOLLITARSQACGH 113
      | | | | | | | | | | | | | | | | : | | | | | | | | | |
Db    240 DPAASLLG-TLOFETQKLRPINAQS---SOTVMNDNRLTIKLVPQLPANNVSGT---- 289

```


Query Match	6.5%	Score 88;	DB 16;	Length 1004;
Best Local Similarity	19.7%;	Pred. No. 10;		
Matches 52;	Conservative 40;	Mismatches 88;	Indels 84;	Gaps 11;
QY	17	LVACSAPIPTNPQVS-----PIKTPSVLIYKDIKIGHHTHE	52	
		::::: :: :	:: :	:
Db	754	LUSSTPUPTTNYLSAFTESLYTDQDSVKRLETLOKDPATTFPSWFSDQLFDPLHP	813	
		::::: :: :	:: :	:
QY	53	HDESVS-----HVGLOAHFETWLOHHIATQEVR	82	
		: :: :	:: :	:
Db	814	NDOQVSPLLSLPKNIHLQTIHGVCCKGVLFMSHPG-EAFPKT--OKTHGFPKDAFE	870	
		: :: :	:: :	:
QY	83	YOAYLOSRLGNLYPPMSOLL-TTARSQAQCHE-----PYQLPPEHLMGQVPVLHLVQ	135	
		::::: :: :	:: :	:
Db	871	LESYLESYLALLQAASHLPKEATILRVPHIEPLIPFPSSPESYL---IRAIHIYE	926	
		::::: :: :	:: :	:
QY	136	DKKSIGI-LPANTQIRSVYRNPELNOCAGAAMSKHLTNSAID-----IWYPDL---EI	185	
		: :: :	:: :	:
Db	927	LLQNOAIPLSQAQAWEYIKTKDTSASOCI-----KKILDSEEDPLTSSFWFHNRDTTEI	980	
		: :: :	:: :	:
QY	186	KSOALYELONRLCQYWLEHENQN	209	
		: :: :	:: :	:
Db	981	CSELNDNVLSQLLSLFINODSQCN	1004	
		: :: :	:: :	:

QY 75 ATKQEVVRYQAYIQSRLGNLYPPM-----SQLTTARSWACGHEPYQLPPEHLWGQI 127



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:15:58 ; Search time 29.76 Seconds
(without alignments)
933.080 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFITTLSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	250	100.0	250	22	AAB60645 Moraxella catarrha

ALIGNMENTS

RESULT 1
AAB60645
ID AAB60645 standard; Protein; 250 AA.
XX
AC AAB60645;

XX DT 04-MAY-2001 (first entry)
XX DE Moraxella catarrhalis strain ATCC43617 BASB120 protein.
XX KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
XX KW genetic immunisation; infection; upper respiratory tract; otitis media;
XX KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
XX KW invasive disease; antibacterial; auditory.
XX OS Moraxella catarrhalis.
XX PN WO200109335-A2.
XX PD 08-FEB-2001.
XX PF 31-JUL-2000; 2000WO-EP07361.
XX PR 03-AUG-1999; 99GB-0018281.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Thonnard J;
XX DR WPI: 2001-159872/16.
XX DR N-PSDB; AAF59797.
XX PT New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
XX PT strain American Type Culture Collection 43617, for use as therapeutic
XX PT agents or vaccines against bacterial infections, e.g. otitis media or
XX PS pneumonia
XX CC Claim 4; Page 64; 75pp; English.
XX CC The invention relates to the Moraxella catarrhalis strain ATCC43617
XX CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
XX CC invention also relates to immunogenic fragments of the BASB120 protein,
XX CC expression vectors and host cells comprising BASB120 nucleic acids, the
XX CC recombinant production of BASB120, vaccine compositions comprising the
XX CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
XX CC compositions comprising the anti-BASB120 antibody, and a method of
XX CC identifying a Moraxella catarrhalis infection via the detection of
XX CC BASB120 proteins or antibodies. The vaccine compositions of the invention
XX CC are useful as prophylactic or therapeutic agents against Moraxella
XX CC catarrhalis infections in mammals, particularly humans. Moraxella
XX CC catarrhalis is a Gram negative bacterium frequently isolated from the
XX CC human upper respiratory tract, which is responsible for several
XX CC pathological conditions. It is responsible for about 15% of otitis media
XX CC cases in children (which can lead to temporary or permanent hearing
XX CC loss). It also causes pneumonia in elderly people, and sinusitis,
XX CC nosocomial infections and, less frequently, invasive diseases. BASB120
XX CC proteins or nucleotides may additionally be used in screening for novel
XX CC antibacterial compounds, and in the diagnosis and staging of infections.
XX CC The present sequence represents the Moraxella catarrhalis strain
XX CC ATCC43617 BASB120 protein.
XX SQ Sequence 250 AA;

Query Match 100.0%; Score 250; DB 22; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.3e-258;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNFNQYFITTLSMLVACSAPIPTNPQVSPKTPSVLLTKDKIGDHHHDESHV 60
DB 1 mknfnqyfitlissmlvacsaaptqpqspktpsvlltkdkigdhthdeshvshv 60
QY 61 GLQAHFETWLOMHATKQEVVRYQAYLQSRGLNGLPPMSQLLTARTSWQACGHPYQLPP 120
DB 61 glqahfetwlmhbatkqevvryqaylqsrnlgnlppmsqlttrswqacghepyqlpp 120
QY 121 EHLNGQIVPTLHLHLDKSRGILLPANTQIRSVYRNPELNQCAGAAKSHLTNSAIDIW 180
DB 121 ehlngqivptlhlhldksrgillpantqirsvyrnpeelnqcagaaakshltnsaidiw 180

Db 121 ehlgqivptlhlyqdlksrgilpantqirsvyrnpelnqcaggaamskhltnsaidiwr 180
Qy 181 PDLEIKSQALYELQNRLCQYWLEHGENONFGLGLYATGAIHLDTOGPRKWGAOFSETNSI 240
Db 181 pdleiksqalyelqnrlicqywlehgengnfglglyatgaihlldtqgfrkwgafsetnsi 240
Qy 241 CRHVLPKNKL 250
Db 241 crhvlpknkl 250

Search completed: July 30, 2002, 16:19:45
Job time: 227 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:17:53 ; Search time 13.06 Seconds
(without alignments)
467.565 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFTTILSSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

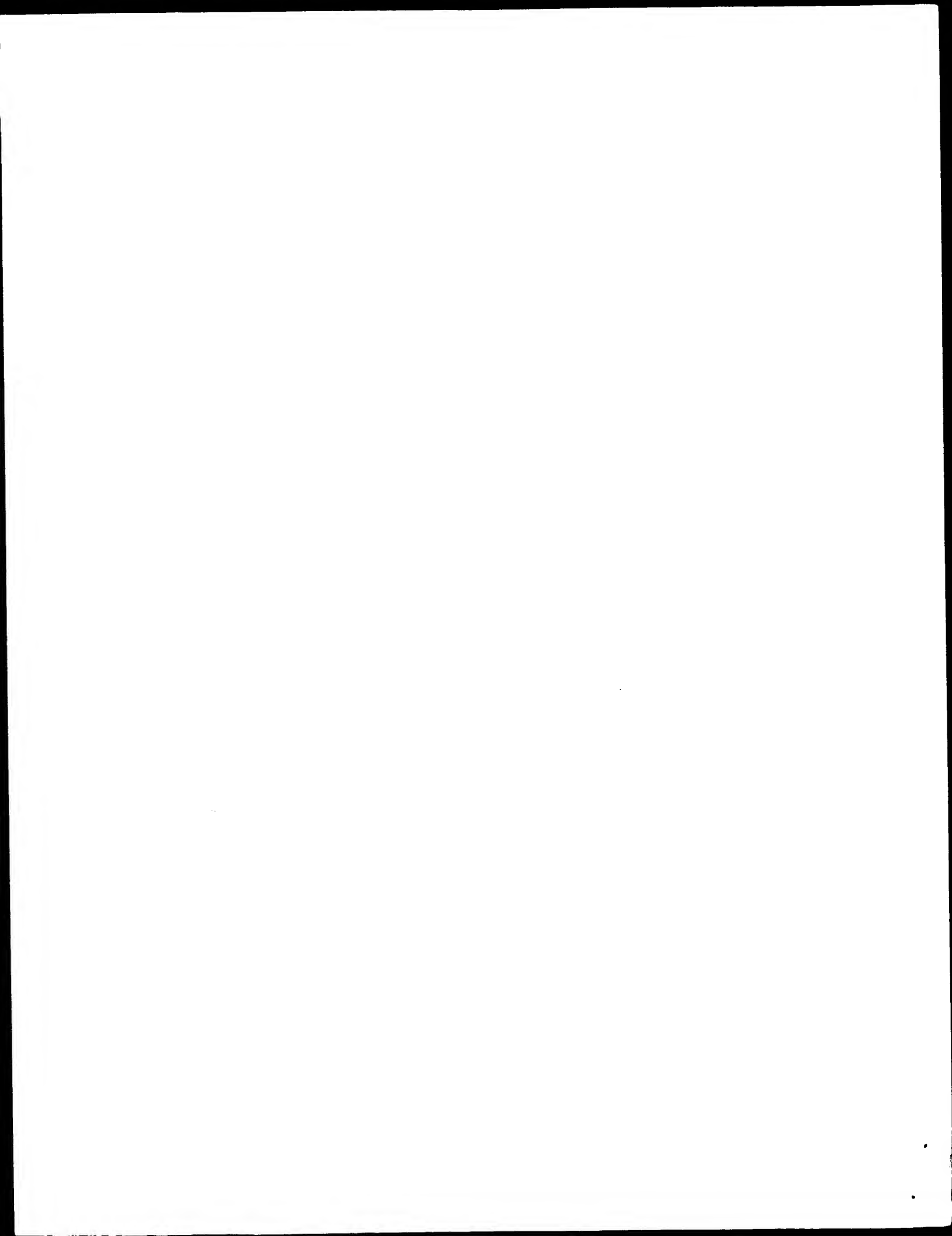
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%	
No.	Score	Match Length DB ID	Description

No matches found			

Search completed: July 30, 2002, 16:20:05
Job time: 132 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:18:23 ; Search time 19.33 Seconds
(without alignments)
1242.749 Million cell updates/sec

Title: US-09-674-779-2

Perfect score: 250

Sequence: 1 MNFNQYFITTLLISSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

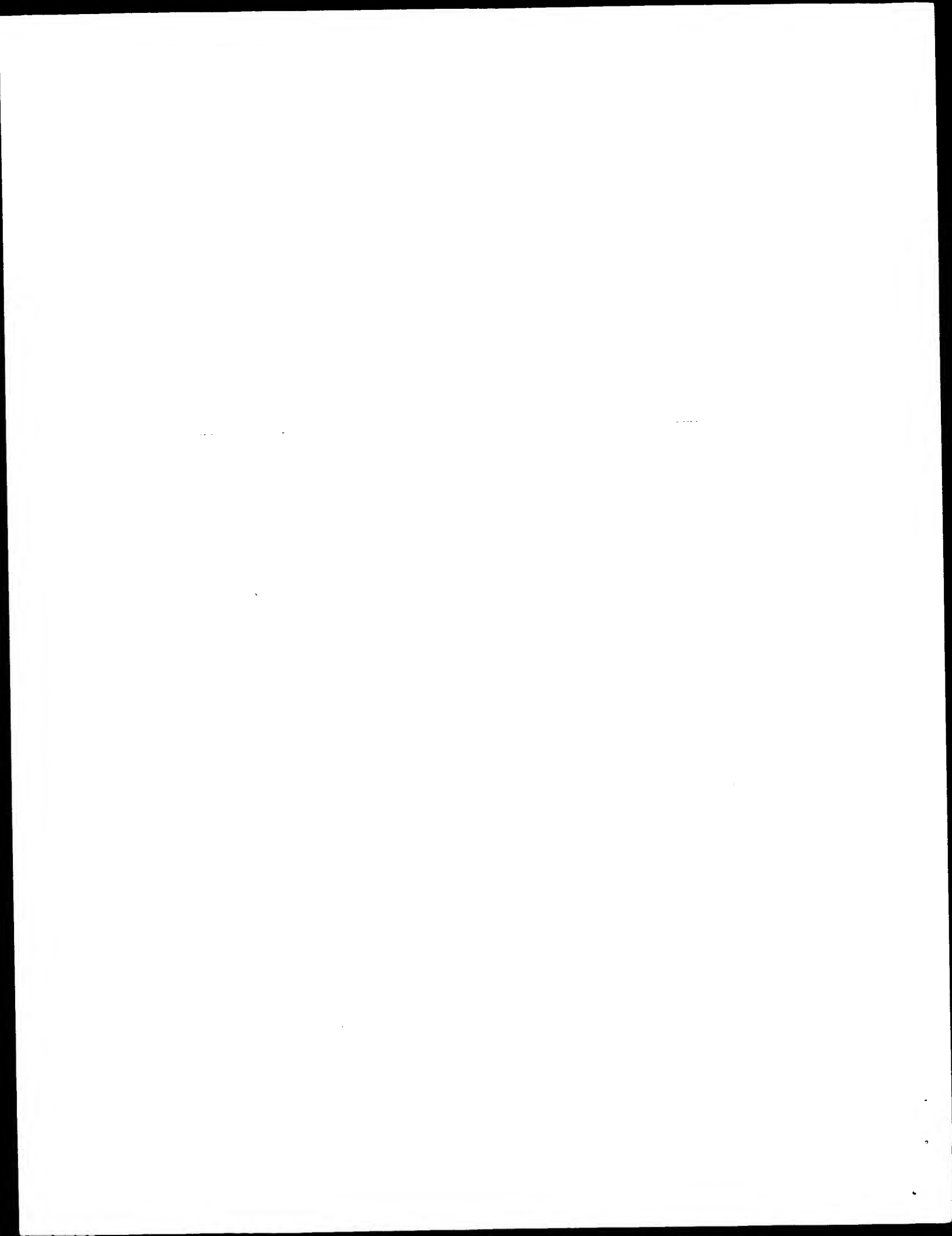
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description
---------------	----------------	-------	--------	----	----	-------------

No matches found

Search completed: July 30, 2002, 16:20:31
Job time: 128 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:20:08 ; Search time 10.97 Seconds
(without alignments)
882.396 Million cell updates/sec

Title: US-09-674-779-2

Perfect score: 250

Sequence: 1 MKNFNQYFITTLSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

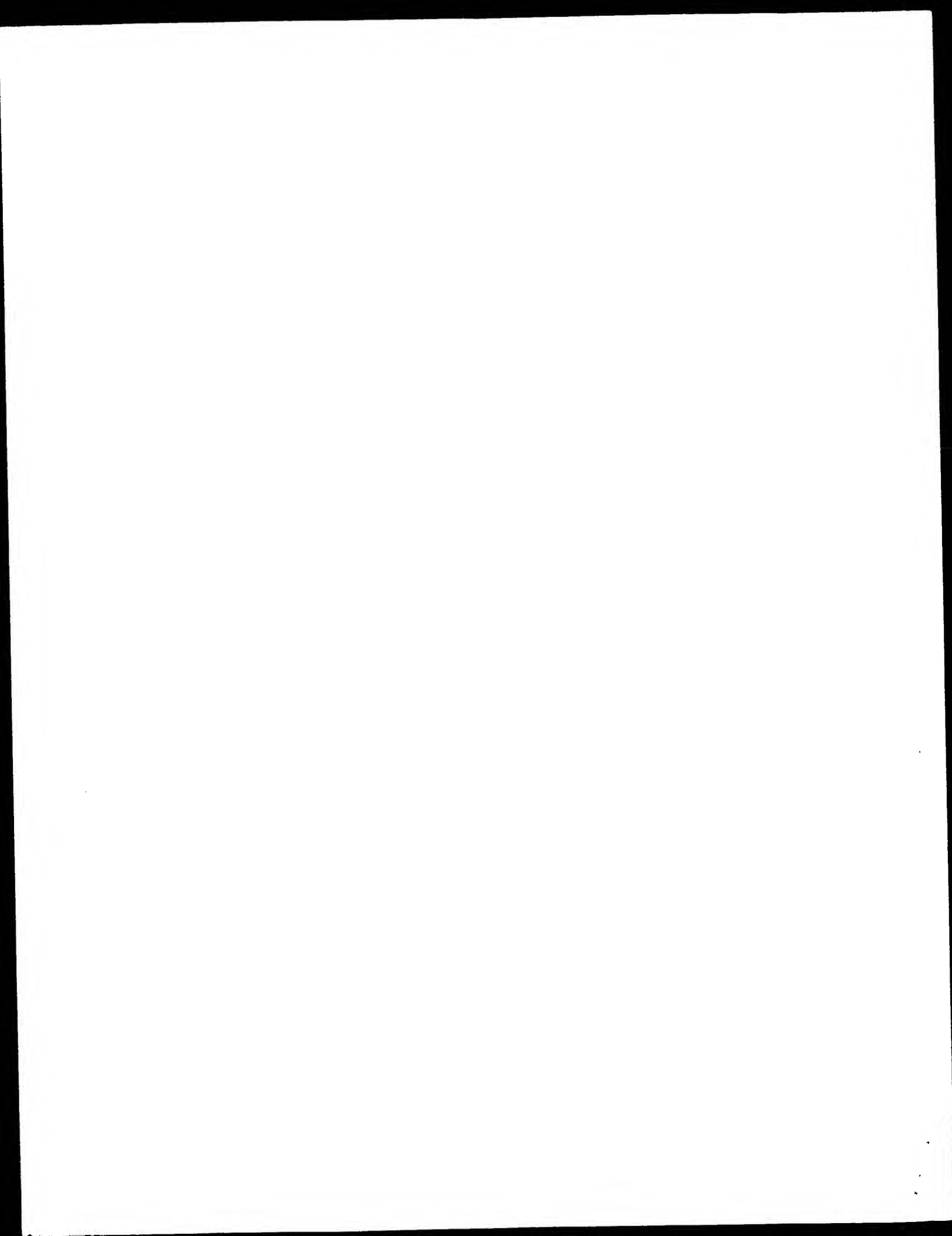
SUMMARIES

Result	Query	%
No.	Score	Match Length DB ID

Description		

No matches found

Search completed: July 30, 2002, 16:23:47
Job time: 219 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:19:48 ; Search time 26.55 seconds
(without alignments)
1628.954 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFITLISMLVAC.....GAQFSETNSICHHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description

No matches found

Search completed: July 30, 2002, 16:23:30
Job time: 222 sec

